

DESARROLLO ECONÓMICO Y DEL EMPLEO A NIVEL LOCAL

La evaluación del desarrollo económico y del empleo a nivel local

Cómo evaluar el desempeño
de los programas y las políticas



ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS

ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS

De conformidad con el artículo 1 de la Convención suscrita en París el 14 de diciembre de 1960, y que entró en vigor el 30 de septiembre de 1961, la Organización para la Cooperación y el Desarrollo Económicos promoverá políticas diseñadas:

- Para lograr la más sólida expansión de la economía y del empleo y a aumentar el nivel de vida en los países miembros, manteniendo la estabilidad financiera y contribuir así al desarrollo de la economía mundial;
- Para contribuir a una sana expansión económica tanto en los países miembros como en los no miembros en vías de desarrollo económico.
- Para contribuir a la expansión del comercio mundial sobre una base multilateral y no discriminatoria, conforme a las obligaciones internacionales.

Los países firmantes de la Convención constitutiva de la OCDE son: Alemania, Austria, Bélgica, Canadá, Dinamarca, España, Estados Unidos de América, Francia, Grecia, Irlanda, Islandia, Italia, Luxemburgo, Noruega, Países Bajos, Portugal, Reino Unido, Suecia, Suiza, Turquía. Los siguientes países se adhirieron posteriormente a esta Convención en las fechas que se indican a continuación: Japón (28 de abril de 1964), Finlandia (28 de enero de 1969), Australia (7 de junio de 1971), Nueva Zelanda (29 de mayo de 1973), México (18 de mayo de 1994), República Checa (21 de diciembre de 1995), Hungría (7 de mayo de 1996), Polonia (22 de noviembre de 1996), Corea (12 de diciembre de 1996), República Eslova (14 de diciembre de 2000). La Comisión de las Comunidades Europeas participa en los trabajos de la OCDE (artículo 13 de la Convención de la OCDE).

Traducido a partir de la versión original en inglés publicada respectivamente con el título:
Evaluating Local Economic and Employment Development:
How to Assess What Works among Programmes and Policies
© 2004 Organización para la Cooperación y el Desarrollo Económicos (OCDE), París.
Derechos Reservados

© 2007 Coordinación General del Programa Nacional de Apoyo para las Empresas de Solidaridad, FONAES, México para esta edición en español.
Publicado por acuerdo con la OCDE, París.

Impreso en México.

La calidad de la traducción al español y su coherencia con el texto original es responsabilidad de la Coordinación General del Programa Nacional de Apoyo para las Empresas de Solidaridad, FONAES, México.

Prefacio

Uno de los mayores retos que enfrentan las autoridades públicas responsables del desarrollo local económico y del empleo —y un reto crítico para los diseñadores de políticas que se enfrentan a todas las formas de desarrollo subnacional— es cómo determinar cuáles programas y cuáles políticas funcionan en realidad. Un corolario a este desafío es identificar, entre todos los programas que sí funcionan, aquellos que proporcionan un mayor valor por el dinero. En un contexto macroeconómico en el cual sólo es probable que las tensiones sobre el gasto público discrecional aumenten, especialmente debido a las implicaciones fiscales o a la transición demográfica, la necesidad de dar respuesta a preguntas sobre la efectividad y la eficiencia de las políticas se volverá más apremiante. De algunos años a la fecha, y en diversos foros, el Programa para el Desarrollo Económico y del Empleo a Nivel Local (LEED, por sus siglas en inglés) de la OCDE ha señalado el déficit en muchos países miembros de la OCDE respecto al volumen y la calidad en la investigación sobre la evaluación de las herramientas utilizadas para mejorar el desarrollo local. Como parte de sus esfuerzos para abordar la escasez de evaluaciones, el Programa LEED organizó una gran conferencia internacional en Viena, en noviembre de 2002, titulada “Evaluating Local Economic and Employment Development” [La evaluación del desarrollo económico y del empleo a nivel local]. Esta conferencia recibió el generoso apoyo financiero y logístico de la Comisión Europea (DG Employment) y del Ministerio de Austria para Asuntos Económicos y Laborales. La conferencia reunió a muchos de los líderes académicos y profesionistas del área de la OCDE preocupados por temas tales como: ¿cómo utilizan los gobiernos los resultados de la investigación sobre evaluación?, ¿cuál es la mejor práctica para evaluar los esquemas que se utilizan con frecuencia para acelerar el desarrollo económico local y el empleo?, y ¿se pueden aplicar métodos rigurosos de evaluación para medir el impacto en localidades enteras de estrategias y programas con múltiples instrumentos?

La evaluación de programas y políticas plantean cuestiones que pueden ser complejas en términos conceptuales y técnicos. Sin embargo se ha hecho un esfuerzo para asegurar que los artículos de investigación sean accesibles para una audiencia no técnica. Los artículos se enfocan a una variedad de programas que tienen su principal impacto en

los mercados laborales locales o en el desarrollo empresarial. Esperamos que estos artículos, y sus implicaciones para la valoración de las políticas precisadas en el capítulo inicial, serán valiosos tanto para la comunidad política como para aquellos encargados de implementar las políticas y los programas.

Son prioridades en curso del Programa LEED: mejorar la práctica de las evaluaciones y avanzar hacia un registro más completo de la evaluación de resultados. Una idea que LEED perseguirá es recopilar un compendio activo en línea de estudios de evaluación de alta calidad. Dicho compendio puede ayudar a ilustrar cómo ciertos retos recurrentes de evaluar han sido abordados en diferentes circunstancias por diferentes instituciones. En conexión con esto, el Programa LEED da la bienvenida al intercambio continuo de visiones sobre todos los aspectos relacionados con la evaluación —y de manera más general a la práctica del desarrollo local— con autoridades municipales, académicos y profesionistas. Además, a finales del 2003, la OCDE estableció, en Trento, Italia, el Centro para el Desarrollo Local. Este Centro tiene una orientación particular hacia los países del Sureste de Europa y tendrá a la evaluación como uno de los componentes centrales de su programa de trabajo.



Sergio Arzeni
Jefe del Programa LEED

Reconocimiento. Esta publicación es el resultado de una tarea en colaboración.

Alistair Nolan y Ging Wong fueron responsables de todos los aspectos sustantivos de la concepción y desarrollo de la Conferencia en Viena de noviembre de 2002, sobre la cual se basaron las contribuciones a este libro. El apoyo esencial para llevar a cabo la conferencia fue proporcionado por Jane Finlay, Jennah Huxley y Sheelagh Delf del Secretariado de la OCDE y por Martina Berger del Ministerio de Austria para Asuntos Económicos y Laborales. Alistair Nolan ha tenido la responsabilidad general de esta publicación, incluyendo la edición de la mayoría de los artículos de investigación. Apoyo decisivo en la realización de este libro fue proporcionado por Sheelagh Delf.

Prólogo

La transparencia y la rendición de cuentas han encabezado la agenda de prioridades de la Administración Pública Federal a lo largo de la década. Este compromiso con la ciudadanía ha llevado a los diseñadores de políticas públicas en México a buscar mecanismos para evaluar con detalle el desempeño de los programas de gobierno e identificar el impacto que tienen en la población.

En este contexto, el Sr. Sergio Arzeni, Director del Centro para la Empresarialidad, PYMES y Desarrollo Local y Director del Programa de Desarrollo Económico y Creación de Empleo a Nivel Local (LEED), invitó al FONAES a realizar la versión castellana del libro: “La Evaluación del Desarrollo Económico y la creación de Empleos a Nivel Local: Cómo evaluar el desempeño de los programas y las políticas”.

La invitación no pudo ser más oportuna y adecuada porque la publicación contiene una compilación de estudios sobre evaluación de programas y políticas públicas, algunos relacionados con el quehacer del propio FONAES, tales como: la evaluación de programas de apoyo al desarrollo de microempresas y pequeños negocios, programas para fomentar la creación de empleo, políticas de focalización, programas de microfinanzas, de capacitación laboral, y de apoyo al tercer sector, entre muchos otros.

Lo anterior, en un momento en que el Gobierno Federal, encabezado por el C. Presidente Felipe Calderón Hinojosa, hace un esfuerzo por realizar una valoración objetiva de los programas para ejercer el gasto público atendiendo a criterios de objetividad, transparencia, eficiencia, economía, eficacia, calidad e impacto social.

Por ello, de acuerdo con el Plan Nacional de Desarrollo 2007-2012 los recursos públicos se asignarán bajo el enfoque de “Presupuesto Basado en Resultados”, circunstancia que plantea un desafío para los servidores públicos, ya que es necesario realizar evaluaciones objetivas, participativas y transparentes de los programas de gobierno que contribuyan a orientar las asignaciones presupuestarias a partir de los resultados y la eficacia de los propios programas.

El Secretario de Economía, Dr. Eduardo Sojo Garza-Aldape, ha puesto especial énfasis en la necesidad de evaluar continuamente el funcionamiento y resultado de las instituciones para elevar sustancialmente la productividad y modernizar procesos específicos del sector público. En este sentido, es necesario conocer el desempeño de los programas y las políticas públicas en términos de cobertura, eficiencia, impacto social y económico, calidad y equidad.

Este libro da luz en esos y otros temas, describe de manera sencilla métodos y técnicas para realizar evaluaciones de impactos y resultados efectivos; pone en la mesa preguntas sustantivas para los diseñadores de políticas públicas, considerando las condiciones específicas del lugar en el que se aplican los programas; identifica dificultades y plantea soluciones a los problemas que puede enfrentar un evaluador; plantea retos y desafíos, describe paradigmas; y recoge el debate académico sobre métodos, técnicas, propuestas y toma de decisiones.

El FONAES identifica en este libro una ventana de oportunidad para que los servidores públicos conozcan y pongan en práctica los hallazgos y metodologías para realizar evaluaciones efectivas y para comenzar con la consolidación de una cultura de evaluación y rendición de cuentas. Espero que este trabajo les sea de utilidad en las tareas que cotidianamente realizan.

Quisiera destacar la permanente interacción que hubo entre el Programa LEED y el FONAES. Este esfuerzo conjunto propició una continua orientación y retroalimentación, en especial con Gabriela Miranda, analista política de dicho programa que participó en los trabajos de revisión de la traducción y cuidado editorial.

De la misma manera, agradezco el apoyo del Centro de la OCDE en México, en especial, el de Annabel Boissonnade y Alejandro Camacho, encargados del área de publicaciones y coordinación editorial. Finalmente, deseo expresar un reconocimiento a todas aquellas personas que directa, o indirectamente estuvieron involucrados con este proyecto, particularmente a Laura Iveth Sánchez Villaseñor, del área de planeación de FONAES, por su participación en la revisión de la traducción al castellano.

C. P. Ángel Sierra Ramírez
Coordinador General del Programa Nacional de
Apoyo para las Empresas de Solidaridad,
Secretaría de Economía,
Ciudad de México, diciembre de 2007.

Índice

Prefacio por Sergio Arzeni	3
Prólogo por Ángel Sierra Ramírez	5
Capítulo 1 Introducción y resumen. La evaluación de programas de desarrollo económico y del empleo a nivel local: un panorama con recomendaciones de políticas públicas por Alistair Nolan y Ging Wong	9
Capítulo 2 Aprendizaje de políticas mediante la evaluación: retos y oportunidades Por Ging Wong	57
Capítulo 3 Evaluación: evidencia para la política pública Por Robert Walker	73
Capítulo 4 La evaluación del impacto de las políticas de desarrollo económico local en los resultados de la economía local: ¿qué se ha hecho y qué se puede hacer? Por Timothy J. Bartik	129
Capítulo 5 Cuatro enfoques para mejorar las prácticas de evaluación en la Unión Europea: Un comentario sobre el artículo de Timothy Bartik “La evaluación del impacto de las políticas de desarrollo económico local en los resultados de la economía local: ¿qué se ha hecho y qué se puede hacer?” Por Daniele Bondonio	161
Capítulo 6 La evaluación de programas dirigidos al desarrollo local y regional: metodología y veinte años de experiencia utilizando el modelo “REMI Policy Insight” Por Frederick Treyz y George I. Treyz	171

Capítulo 7	Comentarios sobre la ponencia de Frederick y George Treyz y el taller “Políticas de análisis para el desarrollo local con modelos de pronóstico” por Robert Wilson	215
Capítulo 8	Evaluación de las políticas de áreas por Brian Robson	223
Capítulo 9	Comentario sobre el artículo y taller “Evaluación de las políticas de áreas” de Brian Robson por Jonathan Potter	249
Capítulo 10	Evaluación de los programas de apoyo a negocios por Eric Oldsman y Kris Hallerg	259
Capítulo 11	Evaluación de los programas de capacitación: impacto en el ámbito local por Randall W. Eberts y Christopher J. O’Leary	283
Capítulo 12	Evaluación de políticas de desarrollo económico local: teoría y práctica por Jeffrey Smith	325
Capítulo 13	La evaluación de programas del tercer sector por Andrea Westall	377
Capítulo 14	Aspectos metodológicos y prácticos para la evaluación de pactos territoriales. La experiencia de Italia por Paola Casavola	399
Capítulo 15	Evaluación de los pactos territoriales para el empleo: aspectos prácticos y metodológicos. La experiencia de Austria por Peter Huber	413
Capítulo 16	Comentario sobre el taller “Evaluación de los pactos territoriales para el empleo: aspectos prácticos y metodológicos” por Hugh Mosley	427
Capítulo 17	Una reseña de las metodologías para la valoración del impacto para los programas de desarrollo microempresarial por Gary Woller	435
Capítulo 18	Un panorama del panel de discusión: evaluando el desarrollo económico local y el empleo por Alice Nakamura	487
	Acerca de los autores y colaboradores	493

Capítulo 1

Introducción y resumen

La evaluación de programas de desarrollo económico y del empleo a nivel local: un panorama con recomendaciones de políticas públicas

por
Alistair Nolan
OCDE

y
Ging Wong
Canadian Heritage y Universidad de Alberta

Los artículos recopilados en este volumen fueron presentados en la conferencia “Evaluating Local Economic and Employment Development” [Evaluación del Desarrollo Económico y del Empleo a Nivel Local], que tuvo lugar en Viena en noviembre de 2002. Esta conferencia fue organizada por el Programa para el Desarrollo Económico y del Empleo a Nivel Local (LEED, por sus siglas en inglés) de la OCDE, con el apoyo financiero y logístico de la Comisión Europea (DG Employment) y el Ministerio de Asuntos Económicos y Laborales austriaco. Lo que motivó la celebración de la conferencia fue la percepción generalizada de que en varios países miembros de la OCDE existe un déficit con respecto al volumen y la calidad en la investigación sobre la evaluación de las políticas y los programas utilizados para mejorar el desarrollo local. ¿Por qué, por ejemplo, es relativamente escasa la bibliografía sobre evaluación del desarrollo local?, ¿es esto resultado de la falta de compromiso público con la práctica de la evaluación en este campo, o es quizá un síntoma de las dificultades conceptuales y metodológicas específicas del desarrollo local? Estos y otros temas fueron explorados en los artículos y discusiones de la conferencia.

La conferencia atrajo a personalidades internacionales líderes en la materia y se propuso abarcar tres aspectos: considerar cómo utilizan los gobiernos la investigación sobre evaluación; examinar cuál es la mejor práctica para evaluar los planes que se utilizan con mayor frecuencia para el desarrollo económico y la creación del empleo a nivel local; y considerar si pueden utilizarse métodos rigurosos de evaluación para medir los impactos de estrategias y programas con diversos instrumentos en una localidad en su conjunto.

Uso y mal uso de la evaluación

Siempre es responsabilidad de los gobiernos asegurar que el dinero público se gaste bien, ya que los usos alternativos del presupuesto compiten constantemente por prioridades en las políticas del gasto. El objetivo de la evaluación es mejorar la toma de decisiones en todos los niveles: en la administración, en las políticas y en las asignaciones presupuestales. La evaluación es objeto de un renovado interés en los países de la OCDE y se reconoce “su importancia en un contexto orientado a resultados, porque proporciona retroalimentación sobre la eficiencia,

la efectividad y el desempeño de las políticas públicas y puede ser crucial para innovar y mejorar las políticas” (OCDE, 1999).

La evaluación trata esencialmente de determinar la efectividad o la aportación del programa, específicamente, el valor agregado de un programa en operación o de una iniciativa pública potencial. Este objetivo primordial ha sido opacado por el hecho de que el trabajo de la evaluación se ha centrado ampliamente en las denominadas actividades formativas* de la evaluación, las cuales proporcionan información para mejorar el diseño y las operaciones del programa. La información exacta a este nivel es relevante pero insuficiente en un régimen administrativo enfocado al ciudadano para el que se generan juicios de valor o excelencia. En este contexto, se da una creciente demanda de evaluaciones del impacto (algunas veces denominadas “acumulativas”) (Canadá, 2000). Estas evaluaciones son intentos sistemáticos de medir los efectos, tanto los pretendidos como los no pretendidos, de algunas intervenciones gubernamentales, o de una mezcla de intervenciones, sobre los resultados deseados. Dichas prácticas de evaluación varían ampliamente en su complejidad y rigor, utilizando con frecuencia análisis comparativos a través del tiempo, entre participantes y no participantes, o a través de estudios de caso detallados. Estas evaluaciones se apoyan típicamente en análisis previos o posteriores a la realización de los programas (“métodos de sistema único”), diseños experimentales y cuasi-experimentales, o análisis detallados de casos individuales cuya aplicación puede ser más factible en contextos prácticos que en grupos de control. Además de la selección del diseño, dichas evaluaciones también requieren tanto de mediciones confiables y válidas, como de procedimientos estadísticos para determinar que tan significativa es una intervención para el resultado que interesa. Al establecer los vínculos entre la política establecida, las decisiones consiguientes y los impactos, la evaluación proporciona un importante mecanismo de aprendizaje y retroalimentación, independientemente del momento específico del proceso de las políticas (State Services Commission, 1999). Incorporar criterios de evaluación a las políticas propuestas obliga a un enfoque *ex ante* de los resultados deseados. La evaluación *ex post* es una importante herramienta para identificar los éxitos y los fracasos de las políticas. En conjunto, las evaluaciones *ex ante* y *ex post* proporcionan las pruebas cruciales que respaldan la rendición de cuentas basada en los resultados.

Sin embargo, la demanda percibida de evaluaciones de calidad de las políticas públicas en general y, en particular, de las de desarrollo local, es baja, dependiendo del país y del periodo en cuestión. Se han dado numerosas explicaciones de ello, relacionadas tanto con la elaboración como con los usos de la evaluación. En

* N.T. Las actividades formativas o la evaluación formativa hacen referencia a los aspectos educativos y de aprendizaje generados como consecuencia de la evaluación.

la vertiente de la elaboración, viene a la mente la referencia hecha por Henry Kissinger al carácter apremiante de la política, cuando *lo urgente roba atención a lo importante*. La evaluación es desplazada por otras demandas inmediatas de los ministros, especialmente en el contexto de escenarios de políticas cambiantes, de los largos periodos de tiempo requeridos para que los resultados sucedan y la necesidad de asignar fondos al desarrollo y mantenimiento de los recursos de evaluación necesarios y las capacidades del personal técnico (ya sea como evaluadores o como clientes inteligentes de las evaluaciones). Las diferentes técnicas de evaluación tienen etiquetados diferentes precios, con la regla de oro de experimentos con muestreo aleatorio de largo plazo en un extremo del espectro, y evaluaciones del proceso en el otro. En los casos en que es forzoso realizar una elección, frecuentemente se hace a favor del enfoque menos costoso. Los aspectos metodológicos también desempeñan un papel en la reticencia de los administradores públicos hacia la evaluación. La evaluación contra resultados se considera simplemente demasiado complicada:

“... la evaluación siempre proporciona respuestas controversiales. Toda evaluación de políticas sociales está imbuida en el problema del contrafactual del cual —nunca se tiene el control—. La experiencia indica que es caro, difícil y controversial” (State Service Commission, 1999).

Hay quienes señalan también las limitaciones de las estimaciones en la evaluación de parámetros implícitos de población o de resultados típicos. Visto desde la óptica práctica de la administración pública y las políticas públicas, una cuestión importante del análisis contemporáneo es el énfasis excesivo en casos promedio. Si bien dicho análisis proporciona gran cantidad de información útil sobre las líneas básicas de las políticas, no suele ser el comportamiento de lo típico e indistinguible lo que más nos preocupa. Las más de las veces es la excepción a la norma estadística —las agencias que representan buenas o excelentes prácticas, la identificación de los programas de alto riesgo, o los programas que pueden cumplir metas múltiples simultáneamente, como eficiencia y equidad— lo que exige reconocimiento o atención para corregirlo (Meier y Gill, 2000). El enfoque en los casos de alto desempeño o de alto riesgo puede poner de relieve las razones que los separan de los casos típicos. El estado actual de la práctica de la evaluación no maneja adecuadamente dichas exigencias de información.

Es una observación recurrente que, según el uso que se le dé, la evaluación es una espada de doble filo. La propia naturaleza de un escrutinio detallado es un vaso medio lleno o medio vacío, ya que hasta los programas ejemplares tienen escollos que pueden plantear al gobierno retos de comunicación políticamente sensibles. Las preocupaciones sobre cómo se podrían usar los resultados de la evaluación, figuran de modo prominente cuando prevalecen actores con intereses propios; de igual modo, ministros y funcionarios pueden estar vinculados con ciertas políticas y ser reticentes a que sufran escrutinio. Muchas veces es

necesario que dirigentes al más alto nivel tengan el valor de resolver este tipo de cuestiones y proteger la función de evaluación y a sus partidarios. Por ello, en diversas jurisdicciones existe el dilema de propiciar un ambiente de transparencia que evite intereses personales y que, al mismo tiempo, detecte los riesgos que surjan al realizar la evaluación. La integridad es el meollo del asunto y el hecho de que se mantenga o no determinará el grado de confianza pública en la función de evaluación y sus usos para mejorar la rendición de cuentas, las prácticas administrativas y la presupuestación.

Desde mediados de los noventa, los esfuerzos sostenidos para modernizar la contraloría realizados por varios países de la OCDE han creado las condiciones necesarias —e indicado mecanismos para ministerios del gobierno— para el desarrollo de la capacidad de evaluar resultados. La nueva práctica de informar sobre el desempeño a los departamentos del tesoro del gobierno exige cada vez más información *ex ante* sobre cómo contribuyen las actividades del programa a la consecución de resultados, así como de información *ex post* sobre cómo se avanza para alcanzarlos. Hoy en día, el proceso presupuestal refleja incentivos para alentar a los ministros a enfocarse en el ajuste continuo de prioridades entre los resultados de las políticas de bajo y alto valor. Al mismo tiempo, el campo de la evaluación está avanzando rápidamente en términos de la creación de rigurosos paneles de datos, en nuevas técnicas y en tecnología de computación. Ante este señalamiento promisorio, dirigiremos ahora nuestra atención al estado del arte en la evaluación del desarrollo local y regional.

Evaluación del desarrollo local

En todos los países miembros de la OCDE se dedica gran cantidad de recursos a programas y políticas para fomentar el desarrollo local y regional. Bartik, en este volumen, describe la magnitud de los gastos para el desarrollo subnacional* en los Estados Unidos. Este autor cita un cálculo anual de 20 a 30 mil millones de dólares destinados por los gobiernos municipales y estatales a programas de apoyo a empresas específicas. El gobierno federal gasta cada año un monto adicional de alrededor de seis mil millones de dólares. Estas asignaciones toman la forma de gasto directo en programas y, preponderantemente, de incentivos fiscales. Además, estas cifras serían considerablemente mayores —y hasta se duplicarían— si se incluyeran más gastos públicos de los gobiernos locales y estatales, y más incentivos fiscales para los negocios. En Inglaterra y Gales, un estudio sobre el gasto de los ayuntamientos locales durante 1999 concluyó que éstos gastaron “322 millones de libras esterlinas en desarrollo económico cada año y que administraron billones de libras de los fondos de regeneración

* N.T. Subnacional se refiere a los estados y municipios.

domésticos y europeos” (Comisión de Auditoría, 1999). Un desembolso público de esa magnitud amerita claramente que se inviertan más esfuerzos en la evaluación de su efectividad y eficiencia.

Especialmente en los lugares más pobres, los programas y políticas para promover el desarrollo económico local van acompañados de intervenciones en una amplia gama de sectores. Las iniciativas incluyen acciones en los mercados inmobiliario, laboral, de información y asesoría empresarial, servicios financieros, de salud, educación y sociales, vigilancia, infraestructura, fiscalización y consolidación institucional de diferentes tipos. Muchos de los programas tienen un largo historial que en algunos casos se remonta a décadas atrás. Las principales características de dichas intervenciones son materia de una abundante bibliografía. Los artículos de este volumen se centran en una serie de programas que tienen su impacto primordial en los mercados laborales locales o en el desarrollo de empresas.

Con el reconocimiento de los vínculos e interacciones entre diferentes intervenciones que tienen el desarrollo local como su punto central, esta obra introduce una desviación respecto a la evaluación tradicional de programas. El enfoque convencional consiste en evaluar instrumentos de las políticas y programas específicos contrastándolos con los objetivos que explícitamente se proponen. De este modo, las evaluaciones de los programas tienden a producir conclusiones aisladas y frecuentemente decepcionantes, sin tomar debidamente en cuenta la interacción y el impacto acumulado de políticas que, adrede o no, funcionan “orientadas por una meta” (Schmid, O’Reilly y Schömann, 1996). Esta perspectiva más amplia reconoce que las políticas diseñadas para influir en el desarrollo de una zona específica no existen aisladamente y que está justificado un enfoque integral.

Hay muy pocas evaluaciones de alta calidad de políticas y programas de desarrollo local

Dada la magnitud de los recursos utilizados para el desarrollo económico y del empleo a nivel local, muy pocos países han hecho una inversión proporcional para generar evidencias, a partir de evaluaciones rigurosas, de qué políticas funcionan y cuáles no. No se trata simplemente de que el financiamiento para la evaluación sea con frecuencia limitado. Adicionalmente, como veremos más adelante, la utilidad de muchos estudios que aparentan ser evaluaciones varía mucho. Por ejemplo, la bibliografía sobre el desarrollo local está repleta de estudios de caso de áreas locales y de sus programas de desarrollo. Sin embargo, muchos de estos estudios describen simplemente un lugar en concreto en un momento determinado. Carecen de datos longitudinales sobre el área o sus residentes, lo cual podría ayudar a detectar las causas de los cambios en

las circunstancias económicas y sociales. En consecuencia, los diseñadores de políticas muchas veces no están seguros de qué políticas elegir como más adecuadas a la circunstancias ni de qué políticas descartar.¹ De hecho, en el mencionado estudio sobre los ayuntamientos locales en Inglaterra y Gales se encontró que en muchos “prevalece la incertidumbre respecto de si sus esfuerzos tuvieron por resultado oportunidades reales para la población de la localidad”.

Al menos en los programas orientados a la localidad, las evaluaciones utilizan frecuentemente métodos y criterios que no son suficientemente estrictos para servir de guía a las políticas. Por ejemplo, hay alrededor de 1 000 incubadoras de negocios en los Estados Unidos. La mayoría recibe algún tipo de financiamiento público y muchas tienen objetivos de desarrollo local o regional. Hay numerosos estudios detallados de planes de incubación. Sin embargo, que los autores sepan, no ha habido aún una evaluación de las incubadoras de negocios en los Estados Unidos que haya utilizado la metodología de grupo de control. A falta de evaluaciones de grupo de control, atribuir los cambios en el desempeño de una empresa a los efectos de la incubación podría ser un error. Asimismo, Boarnet (2001) muestra que, a pesar de la popularidad de las políticas de “zonas” empresariales y de la existencia de un gran número de estudios de evaluación, muy pocos elementos de esta investigación han sido directamente favorables a las políticas. Esto se debe a que los estudios fueron incapaces de identificar qué desarrollos económicos en las áreas objetivo hubiera sido probable que ocurrieran sin la existencia de los programas. Efectivamente, a través de una amplia gama de programas utilizados comúnmente para el desarrollo económico y del empleo a nivel local, se sabe muy poco acerca de los **resultados netos** que se pueden atribuir razonablemente a dichas intervenciones.

Existen otras dificultades que entorpecen el desarrollo de la evaluación. Por ejemplo, muchos estudios tienen tendencia a examinar (el costo de) la producción total en lugar de la consecución de resultados netos. Esto es, los informes se centran frecuentemente en variables como el número de puestos de capacitación provistos, empresas a las que se apoya, unidades de incubación establecidas, volumen de crédito desembolsado, zona de tierra abandonada que se reclamó, etcétera. Por lo general, la compleja cuestión de cómo la oferta de estos productos ha afectado el estatus de los beneficiarios objetivo recibe menos atención. Cuando se reúnen datos pertinentes, suele ser al final de la vida del programa, sin información *ex ante* ni acerca de las líneas básicas. Es más, no es extraño que la evaluación esté elaborada por los mismos patrocinadores del programa que se evalúa.

La creación de empleos es la vara de medición más común de las políticas de desarrollo local. Sin embargo, los estudios utilizan frecuentemente medidas opacas o no estandarizadas de creación de empleos, y esto hace que las aseveraciones del costo por empleo creado no sean confiables e incluso erróneas. Por ejemplo,

si se crea o se traslada una compañía a un lugar determinado y contrata a diez empleados, esto se reporta invariablemente como la creación de diez empleos. Sin embargo, si los diez reclutados simplemente cambiaron de puestos ya existentes puede no haber habido creación de empleos (dicha redistribución podría ocurrir si ha habido puestos de trabajo desplazados de firmas existentes a consecuencia de la competencia de la nueva empresa. Los reclutados podrían también haber abandonado sus empleos anteriores voluntariamente a causa de condiciones más atractivas en los nuevos puestos de trabajo). Normalmente, sólo una parte de las nuevas contrataciones implica creación neta de empleos. Pero también es posible que el bienestar local se hubiera elevado aun cuando todas las contrataciones impliquen redistribución de puestos de trabajo. Por ejemplo, en comparación con los puestos desplazados, los nuevos puestos podrían ofrecer ingresos u otros beneficios a los contratados. A la inversa, el reclutamiento podría estar asociado a una disminución del bienestar social si las personas que son reubicadas de puestos anteriores experimentan pérdidas de ingresos u otras. Los reportes de la creación de empleo no siempre toman en cuenta dichas consideraciones.

Algunos estudios también evalúan con criterios inadecuados. Por ejemplo, los indicadores de creación de empleos pueden ser inadecuados para evaluar los planes de desarrollo empresarial que son uno de los renglones principales de las estrategias locales de desarrollo. Los efectos de dichos esquemas de apoyo empresarial pueden repercutir en una variedad de prácticas empresariales. Pueden repercutir, por ejemplo, en la capacidad de los empresarios para adoptar prácticas administrativas avanzadas, manejar el inventario y el flujo de caja de una compañía, elevar la calidad del producto y disminuir el desperdicio del proceso, ingresar a mercados extranjeros, etcétera. La creación de empleos puede ser un efecto secundario de estas acciones pero no necesariamente surge de manera automática. En términos más generales, muchos programas de desarrollo empresarial elevan el nivel de productividad de la empresa. Esto puede crear presiones para deshacerse de empleados si la demanda de los productos de las firmas es estática. Dichas consideraciones destacan la necesidad de alinear adecuadamente los parámetros de evaluación con la naturaleza de los programas que se evalúan. En el caso del apoyo empresarial, con frecuencia es necesario evaluar los programas en relación con los cambios en las prácticas específicas para el desarrollo de las empresas objetivo, en lugar de evaluar los impactos a corto plazo en la creación de empleos.

La escasez general de evaluaciones de alta calidad es doblemente lamentable en tanto que las iniciativas para el desarrollo local con frecuencia están dirigidas a cumplir funciones piloto o experimentales. De hecho, las políticas piloto son *per se* una de las justificaciones que se esgrimen para los enfoques de desarrollo local. Pero esta función experimental se desperdicia cuando los programas no son debidamente evaluados. Efectivamente, vale la pena hacer notar en relación

con esto que algunas proposiciones fundamentales sobre desarrollo local y las políticas que lo promueven no se comprenden bien. Por ejemplo:

- Existe poca evidencia cuantitativa sobre la aparente eficiencia de los enfoques de desarrollo local *per se*. Hay razones genéricas presumibles para pensar que el diseño y la implementación locales deberían ser prioritarios para algunos tipos de políticas que para otros (por ejemplo, es poco probable la prioridades en el caso de políticas que involucran a importantes economías de escala). Pero es poco frecuente encontrar evidencia cuantitativa de mejoras en la eficiencia económica que procedan de una participación mayor del ámbito local.
- La comprensión de los efectos netos de diversos planes de apoyo empresarial es limitada. Como ya se mencionó, las incubadoras de empresas, a pesar de su proliferación, apenas han sido sometidas a evaluaciones económicas sistemáticas. De modo similar, en países miembros de la OCDE, los programas de microcréditos son escasamente evaluados con técnicas de grupo de control. Y para este reporte únicamente se encontró un estudio sistemático que examinó el impacto local del apoyo al autoempleo [véase Cowling y Hayward (2000)].
- Existe mínima información de cuantificación de costos, beneficios y valores agregados relacionados con las asociaciones locales, una modalidad frecuente de diseño y (en ocasiones) de rendimiento de programas.²
- Como señala Bartik en este volumen, no existe evidencia empírica directa que respalde la idea de que los beneficios del empleo local sean superiores cuando hay un ajuste mayor entre los empleos de reciente creación y las habilidades laborales de los desempleados locales. De manera similar, existe poco respaldo empírico para aseverar que los beneficios del empleo local sean mayores cuando las instituciones del mercado laboral local son más eficientes en la capacitación y la inserción laboral.³

Además —y no sólo para el caso del ámbito del desarrollo local— existen vacíos de comunicación entre los diseñadores de políticas y los profesionales de la evaluación. Por ejemplo, Eisinger (1995) muestra que, desde principios de los noventa, 34 de los 38 estados de los Estados Unidos de Norteamérica han llevado a cabo evaluaciones de al menos una parte de sus programas de desarrollo económico; sin embargo, sólo ocho estados han realizado cambios en los programas a la luz de las recomendaciones de las evaluaciones. La evaluación de la política parece tener poca incidencia en la formulación de la política.

Todo lo anterior no implica que los estudios de calidad sean inexistentes, ni que los profesionales del desarrollo y los diseñadores de las políticas no sean conscientes de los problemas en juego. Los diversos capítulos de este volumen documentan ejemplos valiosos de múltiples evaluaciones de alta calidad. En una primera revisión, Foley (1992) indica una variedad de estudios cuidadosos

que han lidiado con aspectos complejos como el “peso muerto” y los efectos de desplazamiento, de sustitución y multiplicadores.⁴ Particularmente en los Estados Unidos de Norteamérica, se han realizado numerosas y exhaustivas evaluaciones de servicios de extensión científica y tecnológica con base regional [véase Shapira (2003) para una revisión]. Las zonas empresariales han sido analizadas con creciente sofisticación [véase por ejemplo, Bondonio y Engberg (2000) y Boarnet (2001)]. Smith, en este volumen, cita sofisticadas evaluaciones que han abordado problemáticas tan variadas como los efectos de los casinos en el empleo y el impacto en el crecimiento del patrocinio público a los deportes e instalaciones deportivas. Diversos gobiernos nacionales y regionales han creado también lineamientos y capacidad institucional para evaluar programas de desarrollo local (véase por ejemplo, H.M. Treasury, 1995).⁵ Entre las organizaciones homólogas que trabajan con la OCDE, incluyendo agencias de desarrollo regional y local, ministerios del trabajo y otras instituciones públicas, hay un gran interés en la evidencia acerca de lo que funciona y porqué. Con todo, el diagnóstico de la OCDE con respecto a la cantidad y calidad de evaluaciones de políticas pertinentes es de déficit. Y la fuente de evidencia es débil, incluso con respecto a una serie de principios básicos de la práctica de desarrollo local.

¿Por qué es tan escasa la evidencia necesaria de las evaluaciones?

Existen varias explicaciones posibles de porqué la base de evidencia para las políticas es tan débil. Las explicaciones son las siguientes:

- *Posibles objeciones a la evaluación entre los administradores del programa y las agencias que lo implementan.* Estas objeciones pueden surgir a raíz del temor de que se suspenda el apoyo si los programas reciben una evaluación negativa. Éste es un problema *per se* de la evaluación de las políticas públicas. Las objeciones también pueden reflejar el hecho de que las evaluaciones estadísticas más sofisticadas han sido frecuentemente útiles para decidir si una política ha funcionado, pero han sido débiles para describir como podría mejorarse la política (Bartik y Bingham, 1997). Esto puede haber reducido la percepción de la utilidad de la evaluación entre los administradores del programa.
- *Retos prácticos y metodológicos de una evaluación rigurosa.* La medición de efectos de equilibrio general como peso muerto, desplazamiento y sustitución es notablemente difícil, pero la evaluación de políticas de desarrollo local pueden implicar complejidades adicionales dado que los efectos en una zona geográfica se evalúan conjuntamente con los efectos en los grupos objetivo (personas o empresas). Los efectos en los grupos objetivo no se traducen necesariamente en efectos en la zona local. Este es el caso, por ejemplo, cuando los puestos de trabajo que se han creado los cubren inmigrantes, o cuando una mejora en las habilidades de los residentes facilita su colocación laboral, o cuando una mayor

actividad empresarial se traduce en niveles más altos de obtención de insumos fuera del área en cuestión. Tales dificultades se complican aún más si los evaluadores tienen que considerar cómo interactúan varias políticas a través de una zona geográfica que contiene múltiples grupos objetivo. Adicionalmente, algunos resultados del desarrollo local pueden ser difíciles de cuantificar (como la reducción del miedo a la delincuencia en el caso de iniciativas de vigilancia de un barrio, o aspectos del fortalecimiento de capacidades comunitarias).

- *Los costos directos e indirectos de la evaluación.* Los costos directos pueden ser particularmente significativos en evaluaciones con diseños experimentales o cuasi-experimentales. En ocasiones los administradores de los programas también ven la evaluación y el seguimiento como una intrusión de cargas administrativas. Además, las propias evaluaciones pueden involucrar economías de escala (con determinados costos fijos, por ejemplo, en la recolección y la organización de datos) y de alcance (en la medida en que los resultados de una evaluación puedan ser aplicados a otras); dichas economías implican que las evaluaciones sean patrocinadas o emprendidas con frecuencia por los niveles más altos del gobierno. Por lo tanto es poco probable que autoridades municipales produzcan un registro de evaluación sistemática que abarque la amplitud de la innovación de las políticas locales.
- *Incentivos para que las autoridades municipales subinviertan en la generación de conocimientos de evaluación.* Dado que los beneficios de los hallazgos de la evaluación se los pueden atribuir instituciones públicas que no son las que patrocinan los estudios, puede que los organismos locales subinviertan en la evaluación.
- *Sobrecarga de políticas y programas.* Las evaluaciones pueden aparentar ser una inversión poco rentable cuando, como a menudo sucede, las iniciativas del gobierno son numerosas y la población de los programas en activo cambia constantemente (un problema que empeora cuando los programas tienen objetivos múltiples).⁶ El amplio horizonte temporal en el que algunos programas producen efectos mensurables también podría desalentar a los diseñadores de políticas a invertir en la evaluación.⁷
- *La endeble comprensión de las técnicas y los principios de la evaluación, y la falta de evaluadores adecuadamente calificados, de muchas autoridades municipales.* Carecer del personal capacitado necesario para evaluar puede colocar a las autoridades municipales en una posición de desventaja cuando subcontratan estudios de evaluación.
- *En muchos países, la falta de datos apropiados sobre zonas pequeñas.* En algunos casos, las unidades geográficas de las que se recolectan datos —pongamos que de educación o salud— no coinciden con las unidades en las actúan los programas de desarrollo local. En algunos contextos, los datos de zonas pequeñas son simplemente inexistentes.

A pesar de estos desincentivos, algunos países de la OCDE reconocen la importancia crítica de las evaluaciones rigurosas y sistemáticas para tomar decisiones públicas. En Canadá, por ejemplo, es obligatoria la evidencia generada por la evaluación para renovar o reautorizar todos los programas federales. Este requisito de contraloría es un poderoso medio para inducir a invertir en la construcción de una cultura de la evaluación en todo el gobierno federal. Ese compromiso nacional normalmente no ha sido igualado por los gobiernos subnacionales, que son los que operan la mayor parte de los programas de desarrollo local. No obstante, los posibles beneficios de una mayor cantidad y calidad de evaluaciones del desarrollo local podrían ser considerables. Obviamente, las evaluaciones perfeccionadas pueden ayudar a las autoridades locales y centrales a asignar los recursos de manera económicamente eficiente. Las “prácticas óptimas” de diferentes tipos de programas podrían ser estandarizadas de forma significativa, de manera que se pudiera escoger apropiadamente distintas modalidades para la implementación de diversos tipos de programas. Adicionalmente, se observa que la evaluación muchas veces puede ser un ejercicio de aprendizaje para evaluadores y diseñadores de políticas. Dado que el desarrollo local es un esfuerzo multisectorial y que las metas de los programas son algunas veces vagas o incluso inapropiadas, un beneficio potencial importante de intensificar la evaluación podría ser el impulso a las agencias que implementan los programas para especificar claramente cuáles son sus metas.

El seguimiento cuidadoso del programa es también un aspecto delicado y complejo

El seguimiento es una herramienta que puede proporcionar información esencial para administrar el programa y alcanzar mejores resultados. El seguimiento puede también asegurar el contacto cercano con los beneficiarios. Sin embargo, dar seguimiento por lo general no es equivalente a la evaluación, ya que rara vez se ocupa de cuestiones relacionadas con los resultados netos de los programas. Tomando como ejemplo los programas para el mercado laboral, Smith, en este volumen, ilustra cómo los indicadores de desempeño no necesariamente (y por lo general no) sirven como un buen sustituto de las estimaciones de impacto, a menos que exista una relación sistemática entre ambos.

Además de producir información sobre la implementación del programa (y, posiblemente, impactos), los indicadores de desempeño pueden definir relaciones de subcontratación con los proveedores de servicios y servir como instrumento de rendición de cuentas. Asimismo, la elección de indicadores de desempeño para los programas crea incentivos que dan forma a la manera de proveer los servicios. Cuando la continuación del financiamiento de los programas depende de la consecución de objetivos de desempeño pre-especificados, la definición de indicadores de desempeño inapropiados puede tener graves efectos, a veces

difíciles de prever, sobre la implementación y la efectividad del programa. Así que es necesario seleccionar con cuidado los indicadores de desempeño.

Por ejemplo, existe un incentivo para incrementar el rendimiento del cliente si se evalúa el desempeño de un plan de microempresa contrastándolo con el número de personas que entran al programa. Puede alentarse a los clientes a que entren al plan —o aceptarlos cuando deberían ser rechazados— sin importar los resultados probables de la empresa. De un modo similar, cuando las tasas de reembolso de los préstamos se ha utilizado como la principal medida del desempeño en proyectos de microcrédito, el personal utiliza algunas veces procedimientos legítimos de contabilidad para publicar altos índices de reembolso, mientras que los programas han sido diseñados de manera que impiden el reporte de tasas bajas de reembolso (Woolcock, 1999). Igualmente, si los programas son evaluados frente al número de las empresas que ponen en marcha, entonces podría quitarse el apoyo a los servicios que probablemente mejoren la supervivencia de las empresas. Y el financiamiento que se basa en medidas de la productividad, aunque posee la virtud de la simplicidad administrativa, implica hacer pagos después de que los proveedores han incurrido en gastos; esto puede provocar que los apoyos se dirijan a tipos de servicios que requieren poco gasto inicial (Metcalf et al., 2001).

En general, las mediciones del desempeño de una sola variable generan distorsiones en los programas en ejecución que poseen efectos complejos. Debería buscarse indicadores de desempeño que sean capaces de reflejar la complejidad de los programas y de los resultados a los que se da seguimiento. Por ejemplo, usando otra vez el ejemplo de los programas de apoyo empresarial, la medición del desempeño podría combinar datos sobre las empresas de reciente creación —si la creación de empresas es una meta esencial— con indicadores de la supervivencia de los negocios, porque simplemente la creación de empresas no basta. Se podría dar más importancia relativa a proyectos que involucran a empresas en las que la cantidad del capital invertido y los ingresos esperados del negocio son comparativamente altos (en este tipo de empresas es probable que el desplazamiento sea bajo). De manera similar, para evitar el sesgo frecuente de operar con empresarios que habrían sido exitosos incluso sin el programa, podría darse más importancia relativa al establecimiento y la administración exitosa de firmas de individuos que enfrentan dificultades para emprender un negocio (como personas cuyas solicitudes de préstamo han sido rechazadas por el banco).

Metcalf et al. (2001) observa que los proveedores de servicios pueden tener que hacer frente a diferentes requerimientos en los indicadores de desempeño por parte de una serie de patrocinadores de programas. Al respecto, los gobiernos pueden actuar para asegurar la consistencia de los indicadores de desempeño en programas similares. Generar un cierto grado de estandarización en los requerimientos para los datos del desempeño puede reducir la carga administrativa, especialmente para proveedores de servicios que reciben

recursos de más de una fuente. Dicha estandarización puede también mejorar la posibilidad de comparar los programas financiados por el gobierno, lo cual facilitaría la generalización de prácticas óptimas.

También es importante que el seguimiento no se perciba principalmente como un medio de control. Se ha de convencer a los proveedores de servicios de la utilidad de medir el desempeño. En la práctica, el seguimiento se realiza a veces de manera superficial. Muchas veces, los proveedores de servicios no están convencidos de que los datos que se les pide reunir sean útiles.

Panorama de los artículos de investigación recopilados en este volumen

Investigación sobre evaluación y aprendizaje de las políticas

El artículo de Ging Wong, *Aprendizaje de las políticas mediante la evaluación: retos y oportunidades*, reúne variadas y ricas reflexiones del ex director del principal proveedor de servicios de evaluación del gobierno canadiense. El artículo sitúa la función de evaluación en el contexto más amplio de los procesos de elaboración de las políticas, además de proporcionar una cuidadosa exposición de lo que es y no es una evaluación. Traza claras distinciones entre la evaluación y las auditorías financieras, y entre la evaluación y varios sistemas de administración y de rendición de cuentas basados en el desempeño. Wong concluye que las tres principales tareas de la evaluación son: facilitar la rendición pública de cuentas, promover el proceso democrático y mejorar el desarrollo de la investigación y las políticas. Se muestra que el surgimiento de la evaluación en América del Norte, y su posterior adopción en Europa, estuvo directamente relacionado con la necesidad de rendir cuentas de los gastos en el presupuesto gubernamental. Wong identifica también las fuerzas institucionales, nacionales e internacionales que han influido en la expansión de la evaluación en Europa. En relación con esto, observa que el crecimiento de la demanda de evaluaciones parece ser mayor en Europa y a nivel nacional que en los niveles regional y local. Se describen cinco maneras en que la evaluación contribuye al desarrollo de las políticas e implican mejorar:

- El diseño del programa mediante la evaluación de los logros en el desempeño de programas anteriores.
- La implementación del programa a través de evaluaciones del proceso.
- La relación costo-efectividad del programa.
- La administración del programa a través de, por ejemplo, la validación de indicadores y objetivos de desempeño.
- Las capacidades analíticas y de medición.

Se ha demostrado que la evaluación incrementa el conocimiento de las necesidades y los problemas, la efectividad de prácticas y programas, y de la programación.

Wong señala que los usos y la pertinencia de diferentes formas de evaluación —prospectiva, formativa y acumulativa— dependen de la fase de desarrollo de las políticas. Las evaluaciones prospectivas, basadas en compilaciones de datos o en análisis, son particularmente idóneas para etapas tempranas de la formulación de las políticas. Una vez que se ha decidido un curso de acción y se ha establecido el programa, las evaluaciones formativas o de proceso pueden ayudar a identificar y rectificar problemas de la implementación. Por último, las evaluaciones acumulativas buscan aislar los resultados finales atribuibles al programa.

No obstante, el artículo señala que incluso evaluaciones bien preparadas y perceptivas podrían no desempeñar un papel importante en el diseño de las políticas. En la práctica, la evidencia que provee la evaluación constituye sólo un insumo del proceso de desarrollo de la política. Las consideraciones políticas y la opinión pública, por ejemplo, también pueden desempeñar un papel importante. Wong sostiene que los ciclos de desarrollo de las políticas se están volviendo más cortos y que esto constituye una fuente de presión en contra de emprender evaluaciones rigurosas, cuya preparación requiere de más tiempo y dinero. Consecuentemente, hay una mayor confianza en enfoques menos precisos de la evaluación. Un área problemática adicional que destaca el artículo es cómo los evaluadores comunican sus hallazgos al público de una política. Comunicar con claridad el material técnico a personas no especialistas, y que no se dedican a la estadística, requiere de un especial cuidado.

Wong observa que la evaluación es una de las pocas fuentes confiables de prueba de los logros de las políticas y que el compromiso de largo plazo del gobierno con la evaluación es esencial. El reto que Wong observa en el ámbito local es construir estrategias de evaluación participativa que involucren a los numerosos participantes en el desarrollo local. Hay una necesidad paralela y relacionada de estudios de casos locales de alta calidad efectuados de manera multidisciplinaria, que puedan complementar otras formas de evaluación. Dichos estudios de caso deberían servir como insumos para meta-estudios realizados por niveles superiores del gobierno.

Robert Walter proporciona un punto de vista académico y profesional crítico de la *Evaluación: evidencia para la política pública*, apoyándose en estudios de caso recientes de la evaluación de la política del Reino Unido para ilustrar tres desafíos clave. El primero considera las preguntas de evaluación básicas que plantean las políticas públicas y las vincula con técnicas de evaluación para responderlas. La elección de la técnica o modelo de evaluación gira en torno a las preguntas fundamentales de si una política funciona y cómo, y lo mismo sucede con la

perspectiva temporal (pasado, presente y futuro) de la pregunta de la evaluación. Como muestra Walter en su matriz de selección, las técnicas existentes son numerosas y variadas. Para el que patrocina la política, las técnicas reflejan diferentes maneras de realizar las evaluaciones “formativas” y “acumulativas” asociadas con la implementación del proceso y las fases de medición de los efectos directos del ciclo de las políticas. Estos enfoques de la evaluación retrospectiva de las políticas existentes no le pasan inadvertidos a Walter; antes bien, reconoce un creciente interés por las evaluaciones prospectivas para desarrollar políticas potencialmente nuevas. El actual gobierno laborista del Reino Unido, siguiendo la práctica de evaluación de los Estados Unidos, está invirtiendo en el desarrollo de capacidades para evaluar proyectos piloto, prototipos o de demostración antes de tomar decisiones definitivas sobre el diseño de nuevas políticas. Se favorecen los experimentos de muestreo aleatorio y de micro-simulación para comprobar contrafactuales bien definidos, mientras que el meta-análisis se utiliza para agregar y sintetizar resultados de estudios existentes para identificar aspectos logrados de las políticas y la implementación.

Sin embargo, la elección de los instrumentos de evaluación está muy influida por los cambios en el entorno de las políticas. En esto reside la principal contribución de Walter: sus percepciones críticas sobre los contextos institucional y político que han dado forma a los esfuerzos de evaluación pública. La historia británica reciente ha presentado tanto oportunidades como amenazas a la evaluación. Walter ofrece una explicación del giro radical hace veinte años hacia el mayor uso de la evidencia generada por la evaluación para diseñar políticas públicas en el Reino Unido. Esta evolución fue detonada en los ochenta por el surgimiento de la “nueva gestión pública”, que enfatizaba el seguimiento, el control y la medición del desempeño, y que se incorporó a las directivas de supervisión del gasto del Erario Público de Su Majestad (HM Treasury). La creciente demanda de evidencia derivada de la evaluación estuvo acompañada de una práctica sustancial de evaluaciones, aunque siguiendo lineamientos retrospectivos y cuasi-experimentales, en lugar de aplicar la orientación prospectiva y de muestreo aleatorio que se estaba convirtiendo en la corriente principal de evaluación en los Estados Unidos. Para Walter, las razones que explican estas diferencias de enfoque son “principalmente estructurales”. A diferencia de los Estados Unidos, las políticas británicas están altamente centralizadas con una implementación uniforme, ofreciendo pocas variaciones con las cuales probar empíricamente los contrafactuales de las políticas. Además de que las políticas británicas están menos dominadas por la economía positivista que en los Estados Unidos, y las ciencias sociales están más influidas por la investigación tradicional de la acción social que por el análisis cuantitativo y de comparación de grupos, que es el marco dominante en los Estados Unidos y Canadá. El ritmo del trabajo de evaluación se aceleró con la elección del gobierno laborista en 1997 debido a su compromiso de modernizar el diseño de las políticas a través de políticas basadas en evidencias

y mediante “la construcción de una evaluación sistémica de los primeros efectos directos en el proceso de las políticas”. Entre 1997 y 2002, los ministerios del gobierno central iniciaron unas setenta políticas piloto.

Sin embargo, al mismo tiempo Walter aduce que incluso con un fuerte compromiso político de evaluar las políticas, el contexto actual de las políticas británicas no es propicio para apoyar de manera natural el diseño de políticas basadas en evidencias. Ciertos rasgos del mercado de las políticas frustra los diseños y usos más eficaces de las evaluaciones de las políticas, y estos rasgos son: la politización de los hallazgos de la evaluación para abogar por las políticas; la restricción de los cronogramas de la evaluación para acomodar imperativos políticos de corto plazo; la evaluación de problemas políticos que rebasan la competencia de un solo ministerio; la excesiva implementación de programas piloto en todas las localidades, lo cual limita el número de variables de control independientes existentes; y la falta de aprendizaje acumulativo (o amnesia de las políticas) que acompaña a los gobiernos entrantes sin acceso a los expedientes de las políticas creados por sus predecesores.

En pocas palabras, el Reino Unido ha sido bastante exitoso en la integración de evidencias y evaluación de las políticas al proceso de las políticas públicas, pero la sostenibilidad de la práctica actual está muy en duda. Domina la evaluación prototipo a costa de otras estrategias y, hasta la fecha, los resultados sobre impacto y costo-efectividad han sido decepcionantes. Dado que el propio proceso de diseño de las políticas no ha dado cabida a los requisitos para la práctica de una buena evaluación, puede que haya poca motivación en la comunidad académica para ocuparse de dicho trabajo. Para que haya convergencia, es necesario desarrollar una cultura de la evaluación y ésta, según Walter, se puede promover de cinco maneras: debería utilizarse la gama completa de modelos de evaluación para abordar cuestiones relacionadas con cada etapa del ciclo de las políticas; la evidencia producida por la evaluación se debería utilizar de manera apolítica y objetiva; los evaluadores y diseñadores de políticas deberían tener expectativas más realistas de la investigación y de los requerimientos de las políticas; los diseñadores de políticas públicas deberían fijar menores expectativas del impacto innovador de la política; y habría que entender que es excesivamente difícil hacer bien la evaluación, que se requiere de una inversión sostenida, de la acumulación de conocimientos teóricos y prácticos, de datos confiables, de pericia metodológica y de creatividad.

Inventario de la evaluación del desarrollo local

El artículo de investigación de Timothy Bartik —*Evaluación de los impactos en las economías locales: ¿qué se ha hecho y qué se puede hacer?*— examina dos grandes temas: en qué medida se pueden emplear métodos de evaluación que utilizan grupos

de control para evaluar los impactos del programa en localidades completas, y la evaluación de programas para asistir a empresas en comunidades locales.

Bartik llama la atención sobre la necesidad de que los diseñadores de políticas públicas vean más allá de la evaluación de las metas más próximas de los programas. Es necesario evaluar los beneficios públicos de diversos tipos de programas de asistencia. Estos beneficios incluyen los estímulos fiscales y el aumento en el empleo o los ingresos de los desempleados o subempleados. En el ámbito subnacional, la evaluación de estos efectos directos se puede abordar mediante modelos econométricos regionales y de simulación. Este tipo de modelos se toman en consideración en el artículo de Treyz y Treyz, y en el de Wilson. Bartik observa que los beneficios fiscales y en el empleo varían ampliamente, dependiendo de las condiciones demográficas, económicas, del mercado laboral, fiscales y de política social particulares en cada zona del programa. Por ejemplo, los efectos fiscales reflejarán en parte el grado en que los programas afectan la inmigración de población en relación con el crecimiento de empresas. Esto se debe a que en los Estados Unidos, en general las empresas son contribuyentes fiscales netos, mientras que la unidad doméstica promedio usa más servicios públicos de los que paga en contribuciones fiscales.

La exclusión de individuos o zonas enteras del tratamiento de un programa es inherente a los experimentos de muestreo aleatorio; si con fines de evaluación, los programas de desarrollo de zonas excluyeran localidades completas, entonces el programa podría ser polémico políticamente. No obstante, Bartik señala que la asignación de recursos de los programas a localidades en particular se suele guiar por consideraciones políticas más que por consideraciones económicas objetivas. La arbitrariedad que implica la asignación de recursos obvia las objeciones éticas para excluir de la asistencia a ciertos lugares con el propósito de la experimentación aleatoria. Bartik también considera los sesgos presentes cuando se comparan programas de desarrollo de toda una zona. Por ejemplo, los estudios de incentivos fiscales así como de otros incentivos para el desarrollo y la ubicación de las empresas pueden subestimar sistemáticamente los efectos sobre el crecimiento económico local. Esto se debe a que las zonas en las que dichos incentivos son una característica importante de la política tienden a ser las que tienen mayor probabilidad de crecer lentamente, incluso sin los programas de incentivos.

El artículo de Bartik también aborda la evaluación de los programas para apoyar empresas en comunidades pobres. Se hace una distinción entre planes que apoyan a todas las empresas en una localidad dada —como en el caso de algunos programas de zonas empresariales— y los que atienden sólo a un subgrupo de empresas elegibles. Encontrar un grupo de control para el primer tipo de plan está casi automáticamente fuera del alcance de la autoridad local. El artículo de Bartik proporciona también una revisión de las cinco técnicas genéricas existentes para identificar el impacto de un programa a falta de un experimento aleatorio.

Asimismo, se presenta una breve revisión de las evaluaciones de los impactos de los impuestos estatales y locales sobre la ubicación y el crecimiento de las empresas.

El artículo destaca que, mientras que los métodos de encuesta pueden ser valiosos, es más probable que sean confiables cuando evalúan el impacto de los servicios de apoyo que cuando evalúan la asistencia financiera. En este último caso, las empresas pueden estar motivadas a responder de modo que aseguren la continuidad del apoyo económico. De hecho, Bartik señala que como condición para recibir asistencia financiera, algunos programas han requerido incluso que las empresas beneficiarias declaren que la asistencia era crucial para decidir la ubicación o la expansión. En un contexto de este tipo es improbable que las respuestas a encuestas *ex post* sean confiables.

Daniele Bondonio ofrece un comentario sobre el artículo de Bartik y considera las técnicas y los métodos que Bartik analiza en el contexto de los programas financiados por la Unión Europea (UE). Con particular referencia a los programas de la UE para apoyar a empresas en las áreas del Objetivo 2 de la UE, Bondonio argumenta que existe la necesidad de: i) ser más claros sobre qué es realmente una evaluación rigurosa; ii) mejorar la recolección de datos; iii) incorporar mejor las necesidades de evaluación al diseño de las políticas públicas; y iv), sacar provecho de la heterogeneidad implícita en formas regionalmente distintas de diseño e implementación de programas. Bondonio señala que las evaluaciones de los programas copatrocinados por los fondos estructurales de la UE en general sólo intentan medir cambios en las áreas o en las empresas meta; rara vez buscan estimar las diferencias entre los cambios observados y lo que podría haber ocurrido en ausencia del programa. Y aunque modelos económicos regionales como el REMI o el INPLAN* se pueden utilizar para estimar los beneficios fiscales y el empleo en áreas completas, no pueden proporcionar resultados válidos si utilizan indicadores no confiables de los impactos que tienen los programas en dimensiones más próximas de la actividad empresarial. En otras palabras, sin evaluaciones rigurosas del impacto de los programas de apoyo empresarial, los cálculos de efectos multiplicadores más amplios serán imprecisos.

Bondonio hace una serie de observaciones y sugerencias importantes sobre cuestiones de datos. Observa que la necesidad de mejorar los sistemas de seguimiento de programas es ya una muetilla en los estudios de los fondos estructurales de la UE. Sin embargo, existe una necesidad adicional — para fines de una evaluación rigurosa— de obtener datos de alta calidad sobre las áreas y empresas que no han recibido apoyo. La evaluación de programas de incentivos empresariales con una meta espacial podría mejorar si se recolectaran datos a nivel de planta industrial en unidades geográficas pequeñas, sobre todo si estos datos se pudieran combinar con información de los registros de los empleadores

* N.del T. INPLAN: Impact Analysis for Planning Model.

y datos socio-económicos de los residentes. Bondonio señala que en la UE, las regiones NUTS_3 son las unidades geográficas más pequeñas en las que hay estadísticas oficiales y confiables normalmente de fácil acceso; sin embargo, las regiones NUT_3 son más grandes que muchas de las áreas que reciben asistencia —como es el caso de las áreas del Objetivo 2 de la UE— lo cual es un obstáculo para la evaluación con base en una área. La evaluación de los programas pertinentes de la UE se podría facilitar considerablemente mediante la creación de sistemas estadísticos integrales que proporcionen datos de fácil acceso clasificados por unidades geográficas pequeñas que se mantienen estables a través del tiempo. Los sistemas de datos integrados de la UE deberían incluir también registros de las empresas que reciben asistencia de cualquier entidad pública; esto ayudaría a evitar comparaciones entre empresas asistidas por la UE y empresas que simultáneamente reciben apoyo de otras fuentes públicas.

Es necesario también hacer más para incorporar estratégicamente la evaluación al diseño de las políticas públicas. Bondonio indica que podría reasignarse la asistencia a programas a zonas que coinciden exactamente con delimitaciones geográficas sobre las que hay información estadística detallada. También insiste en que las diferencias en la implementación y el diseño de las políticas entre diferentes regiones en las que se aplican los programas del Objetivo 2 brindan la oportunidad de verificar la eficacia de diferentes diseños de políticas.

Uso de modelos de predicción

El artículo de investigación de Frederick y George Treyz —*Evaluación de los programas dirigidos al desarrollo local y regional: metodología y experiencia en la utilización del modelo REMI de percepción de políticas públicas*—* describe el modelo REMI de percepción de políticas públicas, una herramienta para la predicción de la economía regional y el análisis de políticas públicas utilizada ampliamente en los Estados Unidos y en otros países. Este artículo y el debate sobre la evaluación *ex post* están vinculados por el hecho de que los analistas pueden obtener algunos insumos de datos para el modelo REMI a partir de los resultados de las evaluaciones de programas, mientras que la evaluación a nivel micro puede fallar en la captura de importantes efectos más amplios del programa que podrían ser cuantificados utilizando la modelación macro. En términos más generales, la modelación *ex ante* que se ilustra aquí es parte de una búsqueda de evidencia cuantitativa en la toma de decisiones, de la que la evaluación *ex post* es una continuación.

El uso más frecuente del modelo REMI es para cuantificar una extensa gama de impactos económicos regionales/locales que se desprenden de los programas de desarrollo económico (como iniciativas para la atracción de

* Regional Economic Models, Inc. (REMI®) es la empresa proveedora del “REMI Policy Insight®”, modelo para pronosticar y analizar iniciativas políticas.

empresas), inversiones en infraestructura para el transporte y regulaciones ambientales y energéticas. Un beneficio en particular derivado de la utilización del modelo es la identificación de los impactos no previstos del programa o de la política. El modelo integra técnicas de insumo-producto, de cálculo de equilibrio general y de econometría. Las estructuras de insumo-producto rastrean las relaciones entre industrias. Los parámetros de equilibrio general capturan importantes respuestas de largo plazo a señales de precios, costos y salarios. Y las técnicas econométricas validan las bases empíricas del modelo. Los autores brindan una exposición detallada de la estructura del modelo y de los requerimientos de insumo de datos. Utilizando modelos de simulación y apoyándose en la larga experiencia del modelo REMI, los autores aconsejan a los tomadores de decisiones que se aseguren de que las seis características siguientes estén presentes: 1) especificación de las condiciones locales; 2) fundamentos teóricos y estructurales apropiados, 3) métodos integrados de equilibrio general, insumo-producto, econometría y geografía económica; 4) un conjunto de variables de insumo y de producto; 5) resultados año por año; y 6) un registro del uso en una amplia gama de proyectos en regiones diferentes.

Robert Wilson hace una sinopsis del artículo de Treyz y Treyz y de la discusión de la conferencia posterior. Nos recuerda que se han desarrollado otros modelos econométricos macro que se pueden utilizar para ayudar a evaluar intervenciones en el mercado laboral, como el Modelo de Predicción de la Economía Local (LEFM, por sus siglas en inglés) en el Reino Unido. Un tema de discusión importante estaba relacionado con las complejidades y el nivel de experiencia mínimo requerido para utilizar dichos modelos. En efecto, los resultados pueden ser sensibles a cómo se opera con estos modelos y a cuáles son los supuestos que se utilizan. Wilson observa que los modelos se pueden utilizar para proporcionar un contrafactual útil de lo que podría haber sucedido en ausencia de la intervención de la política pública. La conclusión clave de la discusión fue que dichos modelos, como elemento dentro de un enfoque total de evaluación, pueden generar percepciones importantes sobre la evaluación.

Evaluación del desarrollo de áreas específicas en el Reino Unido

El artículo de Brian Robson —*Evaluación de las políticas públicas de áreas en el Reino Unido*— proporciona un panorama omnicompreensivo del desarrollo y la evaluación de iniciativas de regeneración urbana en el Reino Unido (UK) en los últimos treinta años. En ese periodo el Reino Unido ha sido testigo de cambios de énfasis en la regeneración, desde enfoques económicos, sociales, ambientales y basados en la propiedad, hasta el enfoque actual sobre asociaciones y coordinación entre diversos campos de las políticas públicas (como salud, educación y vivienda). Robson describe la principal inversión en evaluación de programas de áreas específicas desde 1997. Tan solo en el ejercicio fiscal 2002-2003,

la Oficina de Representación del Primer Ministro asignó más de ocho millones de libras a evaluaciones de iniciativas de regeneración. Se ha exigido a asociaciones de regeneración locales evaluar y dar seguimiento a sus actividades. Se han publicado en sitios gubernamentales de la red amplias investigaciones sobre evaluación y nuevos datos a nivel de barrios. Y se han adoptado sofisticadas prácticas interdisciplinarias que implican a investigadores y profesionales, así como enfoques de evaluación participativos. En términos de la magnitud de los recursos dedicados y de la sofisticación de las metodologías utilizadas, es obvio que la experiencia del Reino Unido en evaluación de planes con base en zonas contiene lecciones para otros países miembros de la OCDE.

El artículo provee una útil exposición de varios de los retos conceptuales y metodológicos para evaluar planes basados en áreas. La dificultad de estos desafíos es manifiesta por el hecho de que la mayor parte de los estudios todavía se enfoca en el producto en vez de en los efectos directos (netos). Además, quedan sin resolver los problemas de la identificación de interacciones entre los ramales de las políticas públicas. No obstante, se pueden vislumbrar ideas de pasos prácticos a seguir: por ejemplo, Robson cita un estudio de caso que examinó áreas que habían competido por una iniciativa (Reto de la Ciudad) pero que no la lograron. Estas áreas se compararon con otras cuyas ofertas fueron exitosas, obteniendo así inferencias sobre la adicionalidad* del esquema.

Hay un análisis valioso de cómo los aspectos relacionados con los datos restringen la evaluación de áreas. Parece que se ha hecho un nuevo hincapié en la recolección de datos pertinentes a escalas geográficas locales y subdistritales, y a cotejar información procedente de diferentes fuentes. También se han adoptado nuevos enfoques para la recolección de datos que pueden facilitar la evaluación de políticas para zonas específicas. Por ejemplo, Robson describe que el Ministerio de Educación y Habilidades ha decidido reunir datos basados en las direcciones de los alumnos para evaluar así cambios en los logros educativos y relacionarlos con efectos en la zona. Como en otros países, los diferentes ministerios del gobierno muchas veces utilizan distintas unidades geográficas para la recolección de datos y una mayor uniformidad en este sentido podría contribuir a la labor de la evaluación.

Se han hecho importantes observaciones sobre la escasez de evaluaciones que utilizan el análisis longitudinal para rastrear cambios en el estatus del mercado laboral (u otras variables) de unidades domésticas e individuos. Este dato es importante dado que las políticas tienen con frecuencia impactos en una zona que difieren de los impactos originales. Un ejemplo obvio es cuando la creación de nuevos empleos genera vacantes que son cubiertas por inmigrantes. Se hace

* N. del T. Adicionalidad es el grado en el que algo sucede como resultado de una intervención y que no habría sucedido en ausencia de dicha intervención (cf. *Additionality Guide*, Reino Unido, URL: <http://www.englishpartnerships.co.uk/additionality.htm>, consultado el 3 de marzo de 2007).

una breve descripción de los intentos realizados de sondeo longitudinal de este tipo, aunque se señala que el rastreo de los residentes que han abandonado una localidad objetivo puede ser problemático. La necesidad de la evaluación longitudinal se refuerza con la constatación de que algunos de los impactos de regeneración ocurren en horizontes de largo plazo.

Robson comenta que, generalmente, algunos factores de difícil medición son críticos para el éxito del programa. Los ejemplos incluyen la aptitud de la administración, el liderazgo y la sensibilidad a los asuntos de la comunidad. La identificación de dichos factores puede lograrse a través de encuestas y métodos cualitativos, como grupos de discusión. Robson conjetura que la orientación futura de la evaluación en el Reino Unido podría prestar mayor atención a la medición de variables menos tangibles, incluyendo el nebuloso concepto de capital social.

Una impresión persistente que emerge del artículo de Robson es que la dificultad de aislar los efectos de toda un área se exagera con el volumen de iniciativas activas. Parece que un área crucial en la que habría que hacer hincapié es en el diseño de políticas públicas de manera estratégica a fin de eludir algunas de las complejidades de la evaluación. En el artículo no queda claro si en los círculos de las políticas públicas en el Reino Unido se ha pensado seriamente en actuar en esta dirección.

Jonathan Porter sintetiza y en algunos puntos amplía el artículo de Robson y la respectiva discusión en la conferencia. Entre otras observaciones, indica que cuando el apoyo al desarrollo local se da bajo la forma de incentivos fiscales es menos conspicuo y, por lo tanto, menos propenso a la evaluación que en el caso de planes que implican desembolsos presupuestales directos.

Evaluación de los programas de asistencia empresarial

El artículo de Oldsman y Hallberg —*Evaluación de los programas de asistencia empresarial*— ofrece una visión general lúcida de los aspectos implicados en la evaluación de los programas de apoyo a la creación y el desarrollo de empresas. Los programas de apoyo pueden afectar el desempeño de las empresas de diferentes maneras y, por ello, los autores explican que el primer paso esencial de la evaluación consiste en realizar una clara revisión y articulación del diseño del programa. Definir la lógica explícita del programa ayudará a establecer el ámbito de la evaluación, así como los resultados que se evaluarán y las probables relaciones de causalidad. Se hacen importantes observaciones sobre los resultados de los programas y sobre cómo medirlos. Claramente, los efectos directos que se van a medir deben ser los adecuados a las metas del programa. Puede ser que los lectores consideren que esta observación es obvia, pero en la práctica, los planes de desarrollo empresarial suelen estar orientados al logro de resultados —como metas de desarrollo social— que pueden ser totalmente inapropiados. Por ejemplo, en especial en los niveles

municipal y regional de gobierno, los diseñadores de políticas han recurrido a la incubación de empresas como medio para lograr una amplia gama de objetivos económicos y sociales, siendo la creación de empleos la meta más frecuente de los planes de incubación con apoyo público. Sin embargo, contrariamente a la práctica común, la incubación se ha de centrar en el desarrollo empresarial en vez de en el crecimiento del empleo (OCDE, 2003) —el crecimiento del empleo por lo general se desprenderá de resultados comerciales exitosos, mientras que la creación de empleos que surge de la incubación suele ocurrir después de que las empresas arrendatarias ya han pasado por las incubadoras—. Por lo tanto, las mediciones adecuadas del desempeño de la incubadora medirían diferentes dimensiones del desarrollo empresarial. Estas dimensiones podrían incluir, por ejemplo, el tiempo que las empresas necesitan para establecer nichos de mercado o desarrollar nuevos productos, o la adopción de prácticas gerenciales avanzadas. Centrarse en el recuento de puestos de trabajo podría tener por resultado la clasificación de los planes como no exitosos, cuando en realidad éstos podrían producir empleos con el tiempo.

Oldsman y Hallberg observan que, una vez que existe claridad en los resultados a lograr, puede haber más de una medición del resultado logrado. Hay trueques con frecuencia en la elección de diferentes mediciones de un determinado resultado. Las características requeridas de las mediciones de resultados se considera que son: pertinencia, validez, confiabilidad y posibilidad de ponerlas en práctica.

El artículo contiene un panorama muy accesible de las diferentes técnicas de evaluación existentes, con comentarios útiles sobre cómo se pueden aplicar a los programas de apoyo empresarial (se hace la observación de que mientras que las técnicas de evaluación están bien documentadas, ha sido hasta hace poco que se han aplicado a los programas de ayuda empresarial). Se hace la observación poco común de que los estudios que comparan los beneficiarios antes y después del programa en general son de poca utilidad para propósitos de evaluación, pero pueden ser valiosos en ciertas situaciones. Una de estas situaciones podría ser cuando se excluye la identificación de un grupo de control debido a la amplitud de la cobertura de un plan y también cuando hay una relación evidente y fácilmente identificable entre el programa y los resultados en cuestión (se describe un estudio de caso que ilustra esta observación). Los trueques que implica seleccionar la técnica de evaluación —en términos de costos, complejidad y fuerza de la inferencia causal— están también claramente elaborados.

Las recomendaciones de Oldsman y Hallberg hacen eco a varias de las que se presentan en otras partes de este libro. Por ejemplo, hay que tomar en cuenta los requisitos de la evaluación en el diseño del programa. Se han de recolectar los datos básicos y el historial del programa. Hay que emplear múltiples métodos de evaluación. Y las evaluaciones han de intentar demostrar la causalidad, descartando hipótesis alternativas y explicando el mecanismo causal. Uno de

los pocos aspectos que no reciben atención en este artículo se refiere al momento oportuno para la evaluación. Es importante decidir cuándo deben realizarse las evaluaciones porque diferentes programas de apoyo empresarial tienen diferentes periodos de gestación. Algunos ofrecen la posibilidad de beneficios casi inmediatos. Por ejemplo, una incubadora de empresas podría proporcionar acceso instantáneo a bienes inmuebles a las firmas elegibles. En otros programas podrían ser necesarios meses o años antes de que el cambio sea evidente. En particular, el impacto en la creación de empleos de los programas de apoyo al empresariado tiene tendencia a no percibirse en el corto plazo. El periodo de tiempo sobre el que se realiza una evaluación podría exceder el de la duración de la participación en el plan, dependiendo del tipo de programa que se esté examinado.

Podrían hacerse una serie de puntualizaciones adicionales sobre los impactos de las políticas públicas para fomentar el empresariado en el desarrollo local. La primera es que, por sí mismos, los programas de desarrollo del empresariado son con frecuencia demasiado pequeños con relación a la economía local para haber registrado efectos significativos (Bartik y Bingham, 1997). Además, como se mencionó en artículos anteriores, las evaluaciones que atribuyen exitosamente resultados a un programa de apoyo al carácter empresarial son sólo un primer paso en la evaluación del impacto en la economía local. Los evaluadores necesitan ir más allá de los efectos inmediatos del programa y rastrear sus efectos multiplicadores en el ingreso y el empleo, así como sus consecuencias fiscales. Además, podría ser necesario tomar en cuenta en la lógica de la evaluación el hecho de que los impactos específicamente locales pueden estar mediados por desarrollos económicos independientes de los programas de fomento del empresariado, como la inmigración laboral. Por ejemplo, para determinar el impacto de un plan de apoyo a la empresa sobre el empleo local es necesario seguir varios pasos (véase Storey, 1990). Por ejemplo, tendría que establecerse:

1. Si en el momento de la evaluación se ha logrado todo el impacto del empleo (esto puede ser difícil de valorar, incluso para las propias firmas).
2. ¿Cuántos de los empleos resultantes del Paso 1 son empleos que no se hubieran generado sin el programa?
3. ¿Cuántos de los empleos resultantes del Paso 2 no implican el desplazamiento de empleos hacia otras empresas?
4. ¿Cuántos de los empleos identificados en el Paso 3 proporcionan trabajo a residentes o a inmigrantes? (adicionalmente, muchas veces los diseñadores de políticas querrán saber con respecto a los residentes empleados identificados en el Paso 4, ¿cuántos salieron del desempleo o cuántos provienen de grupos objetivo en particular, como la juventud o las minorías étnicas? A los diseñadores de políticas también les podría interesar si estos trabajos son de corta o larga duración y de alta o baja calidad).

Evaluación de programas para el mercado laboral

En los Estados Unidos, los economistas de la fuerza de trabajo están muy implicados en la evaluación de programas públicos de empleo y capacitación desde los sesenta, cuando el presidente Johnson renovó la capacitación laboral federal como un elemento esencial de la *Lucha contra la Pobreza* de su administración. Además, la insistencia estadounidense en evaluar estos programas del mercado laboral es claramente diferente de la práctica europea usual. Por consiguiente, este volumen comparativo de la OCDE estaría notablemente incompleto sin una valoración del papel que estas evaluaciones han desempeñado en el desarrollo de innovaciones técnicas y en dar forma a la política pública del empleo en los Estados Unidos. Para ello, recurrimos a Randall Eberts y Christopher O'Leary. En su artículo —*Evaluación de programas de capacitación: los impactos a nivel local*—, estos autores examinan la evolución de las características clave de las políticas de capacitación estadounidenses en los últimos cuarenta años y documentan las evaluaciones relacionadas con ellas, a la vez que comparan y contrastan las prácticas de evaluación federales y estatales.

Los principales hallazgos se pueden resumir como sigue. Primero, mientras que la mayoría de los programas gubernamentales regularmente se evalúan para determinar la renovación o la reautorización del programa o el diseño de nuevas iniciativas de capacitación, los programas que apoyan los estados han sido objeto de menos escrutinio. Las razones que explican esta brecha son: a) la complejidad adicional de evaluar programas estatales y locales que tienen comúnmente una multiplicidad de objetivos, ejemplo primordial de ello es la capacitación a la medida de las necesidades de las empresas como una manera de promover el desarrollo económico local y regional; b) presiones para abogar por las políticas que disminuyen el compromiso político con la evaluación; y c) restricciones presupuestales. En cambio, Eberts y O'Leary señalan que desde los noventa, las evaluaciones se han vuelto parte integral de muchas de las iniciativas de reforma de la seguridad social a nivel estatal y municipal. Esto ha sido resultado del acuerdo federal de conceder a los estados que renuncian a este derecho de cambiar aspectos de la programación federal, con la condición de que evalúen los programas sociales. Gran parte de lo que se sabe en la actualidad sobre los efectos de los incentivos de las diferentes reformas de la seguridad social proviene directamente de estas evaluaciones locales.

Segundo, las técnicas más populares de evaluación de la capacitación laboral son el seguimiento del desempeño, que rastrea los resultados brutos, y la estimación del impacto neto, que evalúa el valor agregado de una intervención. Las evaluaciones de la capacitación en los Estados Unidos se contemplan esencialmente a través del lente de las evaluaciones del impacto neto, que comparan los resultados promedio de una muestra representativa de participantes en el programa con una muestra similar de no participantes. El diseño de la

evaluación se hace normalmente mediante asignación aleatoria de los individuos o mediante una comparación de una selección de muestras basada en características observables en un cuasi-experimento. En el caso de la capacitación laboral, dichas evaluaciones también se consideran y ajustan mediante una sobrestimación de los impactos netos del programa debidos al efecto denominado “descenso de Ashenfelter” observado en los ingresos anteriores a la separación del trabajo en el caso de trabajadores desplazados. Se describen siete criterios para obtener buenos métodos de estimación y se enumeran por orden de importancia. Se ofrece una discusión más a fondo sobre los aspectos prácticos de evaluar la capacitación laboral, incluyendo tamaños de muestras, selección del lugar, selección de la muestra, implementación de la encuesta y preparación de los datos para el análisis. Después se presentan resultados clave de la evaluación sobre qué es lo que funciona, para quién y por cuánto, en la serie de políticas para el empleo implementadas en Estados Unidos, empezando por el Decreto de Capacitación para el Desarrollo de la Mano de Obra (MDTA, por sus siglas en inglés) de 1962, el Decreto Integral del Empleo y la Capacitación (CETA) de 1973, el Decreto para la Asociación de Capacitación Laboral (JTPA) de 1982, y el Decreto para Inversión en la Fuerza de Trabajo (WIA) de 1998.

Tercero, un gran número de estados subsidian la capacitación de los trabajadores mediante la oferta de programas de capacitación personalizados a las necesidades de las empresas locales, no con el fin principal de mejorar la posibilidad de empleo a los trabajadores, como en los programas federales, sino como un medio para promover la ubicación y retención de empresas. En consecuencia, las evaluaciones estatales y municipales son tareas complejas que, en términos ideales, combinan los beneficios netos de la capacitación para el trabajador individual con el efecto neto de la capacitación sobre la productividad y los salarios y, adicionalmente, amplios indicadores regionales como la creación de empleos y la reducción de la pobreza. Todo esto introduce complejidades e imprecisiones adicionales a la evaluación. Para minimizar estas dificultades, Eberts y O’Leary identifican tres tipos básicos de metodologías de evaluación que pueden ser apropiadas para evaluar los programas estatales de capacitación personalizados: construcción de grupos de comparación de empresas y no de individuos; utilización de modelos econométricos para evaluar explícitamente las relaciones clave entre la intervención y los resultados deseados; utilización del enfoque de una empresa representativa para estimar los efectos relativos de programas en el estatus financiero de la empresa. Se proporcionan ejemplos para ilustrar estas estrategias de evaluación alternativas. La conclusión general es que la evidencia sobre programas de capacitación laboral financiados por los estados es muy limitada y, por lo mismo, nuestra comprensión de la eficacia de la capacitación en el ámbito local es irregular e imprecisa. Dicho eso, hay evidencia en un puñado de estudios de programas específicos de capacitación estatal que indica que estos subsidios a la empresa privada aumentan la capacitación y mejoran la productividad, tanto de

las empresas como de los trabajadores, a la vez que aumentan los salarios, mejoran el ambiente de trabajo y expanden la fuerza de trabajo.

Jeffrey Smith, en su artículo titulado: *Evaluación de políticas locales de desarrollo económico: Teorías y práctica*, proporciona un punto de partida complementario pero diferente a las cuestiones metodológicas destacadas por Eberts-O'Leary y Bartik. Argumenta que la bibliografía académica actual sobre cómo evaluar programas ha tenido avances significativos en los últimos quince años, pero que la práctica de la evaluación sigue estancada en los setenta. Proporcionando una guía práctica y relativamente no técnica, se propone hacer más asequible la bibliografía académica.

El plan de exposición de Smith se centra en cinco temas claves para la evaluación de los programas de desarrollo económico local: a) la elección de estimadores econométricos para la evaluación; b) cómo los impactos diferenciales del programa en individuos, grupos, firmas y localidades afectan tanto la práctica de la evaluación como nuestra capacidad de pensar el diseño de la evaluación y su interpretación; c) la preocupación por las implicaciones que tienen los efectos del equilibrio general en la evaluación de las políticas; d) cuándo no hacer una evaluación, particularmente en programas pequeños; y por último, e) la evaluación como un método para asegurar que los operadores del programa estén al servicio de los intereses de los contribuyentes y del bien público, en lugar de simplemente canalizar dinero a las partes interesadas políticamente influyentes, con la cobertura de que justifican el desarrollo económico local.

Smith llama la atención en concreto sobre percepciones procedentes de la bibliografía para avanzar en los métodos y en la práctica de la evaluación. En primer lugar, no existe una varita mágica. Cada categoría de estimadores econométricos proporciona la respuesta correcta sólo bajo ciertas combinaciones de datos existentes, considerando cómo tiene lugar la participación del programa, y determinados parámetros de interés. En cambio, es esencial elegir e interpretar estimadores econométricos de la evaluación para tener un plan de evaluación que muestre un mapa de datos, instituciones y determinado parámetros de interés y, al mismo tiempo, tome en cuenta los efectos heterogéneos del tratamiento. Esto garantizaría que la selección de los estimadores se apega a reglas estrictas. Segundo, dado que los programas de desarrollo local en general tienen por objeto crear efectos de equilibrio general o multiplicadores, la elección del estimador que se va a utilizar y la unidad de análisis para la evaluación fallará o estará sesgada por los efectos de equilibrio general.

Evaluación de programas del tercer sector

El artículo de investigación de Andrea Westall —*Evaluación y programas del tercer sector*— revisa los enfoques que han surgido para la evaluación de

programas implementados por organizaciones de servicio sin fines de lucro y por las llamadas empresas sociales.⁸ Dichas organizaciones del “tercer sector” han ido desempeñando un papel cada vez más prominente en las iniciativas de desarrollo local. Su función más conocida tal vez sea como institución intermediaria en el mercado laboral que brinda experiencia laboral y capacitación con el fin de facilitar transiciones para trabajar en el mercado laboral. La evaluación del tercer sector es importante debido al creciente compromiso público con dichas organizaciones. También es importante dado que dichos organismos hasta ahora han sido objeto de una escasa evaluación cuidadosa, pese a que desempeñan funciones que, en algunos casos, han dado respuestas innovadoras a problemas locales económicos y sociales. Además, las propias organizaciones del tercer sector desean medir con precisión la gama de beneficios que generan y con ello mejorar el suministro de sus servicios.

Las organizaciones del tercer sector con frecuencia están acostumbradas a dar seguimiento y a reportar insumos, como horas voluntarias de trabajo y los montos y las fuentes de aportaciones financieras. Muchas recolectan información sobre efectos directos en los clientes. Pero hay un relativo desconocimiento de técnicas más sofisticadas para la valoración del impacto, como valoraciones confrontando datos pre- y post-programa, procedimientos para muestreo de clientes, seguimiento a clientes que abandonan el programa, diferenciación demográfica de los datos de beneficiarios y el uso de grupos de comparación.

Westall sostiene que la subevaluación de los programas del tercer sector refleja en parte las restricciones de su capacidad interna, así como los tipos de criterios de seguimiento y evaluación (basados en el producto) utilizados por los patrocinadores del programa. También se dice que la diversidad de los requisitos para los reportes a los múltiples patrocinadores del programa, ha obstaculizado la evaluación. En parte, el carácter superficial de algunas evaluaciones podría reflejar el hecho de que este tipo de programas rara vez están financiados por autoridades del ámbito federal, que son las que usualmente proponen formas rigurosas de evaluaciones. Además, como los volúmenes de financiamiento a los planes del tercer sector son pequeños, pudiera no valer la pena utilizar técnicas de evaluación más rigurosas y costosas.

La existencia de limitaciones en la capacidad interna de evaluación es en parte un problema de financiamiento. Esto acentúa la necesidad de etiquetar recursos para la evaluación en los presupuestos públicos y asignarlos a organismos del tercer sector; dichos recursos podrían ayudar a solventar gastos en tiempo del personal, correspondencia, encuestas telefónicas, etcétera. Las organizaciones nacionales y municipales públicas y las no gubernamentales podrían también organizar la asistencia técnica para mejorar las capacidades de evaluación en los organismos del tercer sector.

Son de importancia considerable las ventajas potenciales de involucrar a miembros de la comunidad y a los beneficiarios en el proceso de evaluación. Westall sostiene que la participación en la evaluación es importante, como medio de recoger información y como una forma de crear el sentido de pertenencia al programa. Se reconocen potenciales desventajas, pero quizás no se les da suficiente peso. Westall observa que entre las desventajas están los recursos que se requieren para organizar la participación y el peligro de obtener respuestas sesgadas si predominan los puntos de vista de los residentes con más voz e influencia. No obstante, en términos generales existe el peligro de que los enfoques participativos (si la participación implica algo más que simplemente ser un encuestado) puedan entrar en conflicto con la objetividad que debe ser el fundamento de la evaluación. Además, en este volumen Smith observa que existe poca evidencia directa de la confiabilidad de la autoevaluación de los participantes. Él se refiere a los hallazgos a partir de evaluaciones experimentales y de la teoría de la decisión conductista, que sugieren que la mayoría de las personas carecen de competencia para contestar la clase de preguntas contrafactuales que los evaluadores necesitan plantear. El artículo de Woller menciona una serie de debilidades adicionales en los enfoques participativos. Éstas incluyen una falta de estandarización de las técnicas que se utilizan (aunque éste no es necesariamente el caso si los organismos centrales se esfuerzan en garantizar procedimientos y formatos en común), y los problemas que surgen de las interpretaciones dispares del impacto dentro del mismo grupo meta. Como observa Woller, no se puede afirmar que las técnicas participativas establezcan la causalidad. Pero a través de metódicas verificaciones cruzadas de diferentes formas de evidencia se puede proponer, al menos, un caso creíble de causalidad. La insistencia de Westall en la combinación de metodologías —técnicas más rigurosas con encuestas (participativas)— se reitera en varios de los artículos recopilados en este libro.

Evaluación de pactos territoriales de empleo

Las experiencias prácticas de evaluación de pactos territoriales de empleo por los ministerios del gobierno fueron el tema de una sesión de trabajo. Este taller incluyó dos artículos —*Evaluación de pactos territoriales de empleo en Italia* de Paolo Casavola y *Evaluación de pactos territoriales de empleo en Austria* de Peter Huber— así como un resumen y un comentario sobre la discusión en *Evaluación de pactos de empleo locales en Austria e Italia* de Hugh Mosley.

Casavola introduce el concepto de pacto territorial de empleo como “un instrumento específico de las políticas públicas dirigido a promover el desarrollo local a través de incentivos financieros a un grupo de proyectos con base local e integrales, diseñados por una coalición de actores locales (privados y públicos)”. Aunque los pactos territoriales individuales no cubren grandes zonas, se utilizan ampliamente en Italia, donde hasta septiembre de 2002 operaban 230. La

mayoría de ellos fueron seleccionados mediante un procedimiento nacional y unos pocos fueron seleccionados mediante un procedimiento establecido por la Comisión Europea. En teoría, dichos pactos territoriales pueden acelerar la ubicación de actividades de valor agregado mediante la promoción de economías de aglomeración y otras sinergias. Para evaluar estas políticas, Casavola distingue dos cuestiones separadas pero relacionadas: mecanismos para inducir el desarrollo local (selección de proyectos y capacidad de coalición pública-privada) y resultados económicos finales. En el caso italiano, la evaluación formal —tanto formativa como acumulativa— no está determinada para abordar la información y los requisitos necesarios para la rendición de cuentas. Esto refleja en parte la opinión de la administración de que los pactos deberían estar en operación durante el tiempo suficiente con el fin de evaluar los resultados. Provisionalmente, se ha establecido el seguimiento administrativo de las contribuciones financieras a los pactos. También se han comisionado estudios académicos y trabajo de campo para elaborar estadísticas territoriales asociadas con la cobertura del pacto y para llevar a cabo encuestas de los empresarios que son partes interesadas y de informantes clave. Con base en estas bases de datos, Casavola reporta hallazgos preliminares de la evaluación sobre la regresión de mínimos cuadrados ordinarios (MCO) del desempeño del pacto en proceso (índice de eficacia del gasto de la contribución financiera pública) y variables explicativas por tipo de pacto, condiciones económicas iniciales y capital social. La participación de las partes interesadas parece ser un factor significativo para una disposición e implementación más rápidas.

Huber esboza el enfoque que se ha elegido para evaluar los pactos territoriales del empleo en Austria. Desde la introducción realizada por la Comisión Europea en 1997 de cuatro pactos austriacos cuyo propósito era combatir el desempleo, se han creado nueve pactos adicionales con base en provincias a través de un plan nacional de acción para apoyar este tipo de acuerdos territoriales mediante subsidios específicos. El propósito de estos pactos provinciales varía, desde la coordinación del mercado laboral provincial y medidas económicas presupuestales, hasta la coordinación de innovaciones en las políticas. Se analizan los primeros estudios y evaluaciones externas de la estructura y el desarrollo de los pactos, así como los esfuerzos de autoevaluación. Un problema particular de la evaluación es la escasez de datos administrativos significativos asociados con actividades de coordinación. Sin embargo, los pactos generan reportes administrativos regulares, que pueden ser una fuente de información para generar indicadores sobre el desarrollo del pacto. Dadas las circunstancias, Huber recomienda una estrategia de evaluación que se centra directamente en los procesos: una evaluación formativa que evalúa la formulación de las políticas públicas, su implementación y captación. Esta orientación también insiste en determinar si los pactos han contribuido a establecer capital social en la región. Ésta sigue siendo una pregunta empírica que la evaluación formativa ha de responder.

Mosley pone en perspectiva el problema de la evaluación cuando nos recuerda que las redes de los pactos no son organizaciones formales con sus propios presupuestos y empleados, sino que son asociaciones voluntarias en las que los miembros que las constituyen mantienen sus propias identidades. Por lo tanto, un valor de dichas redes es potenciar una cantidad de actividades conjuntas desproporcionadamente alta que de otro modo no existiría la clave de la evaluación o el contrafactual que hay que examinar. Las supuestas fuentes de valor agregado son efectos de la sinergia y la formación de capital social, asociados con impactos favorables a largo plazo en el empleo y el desarrollo económico locales. Para responder a la pregunta “¿cuál es la diferencia cuando se evalúan pactos territoriales?”, Mosley proporciona un útil punto de partida de cómo pensar en los requerimientos de la evaluación para todas las políticas de desarrollo local. En primera instancia, y en contraste con las políticas públicas orientadas a individuos o a empresas, la pregunta central de la evaluación es si se ha logrado el efecto local o de zona. Una segunda pregunta de la evaluación es el efecto de los acuerdos de gobierno sobre el nivel y la mezcla de las políticas públicas locales. Otra pregunta característica de la evaluación es el impacto de los cambios en el gobierno inducidos por el programa sobre los efectos directos agregados de las políticas en las regiones. El análisis de grupo de control/tratamiento es inadecuado en este caso y se indica en cambio un análisis alternativo de impacto agregado que podría utilizar un marco de series temporales o transversales, o una comparación por emparejamiento de las unidades regionales. La pregunta final y clásica de la evaluación es el impacto del programa sobre los participantes individuales. Dada la heterogeneidad de los pactos territoriales y la hipótesis de que la implementación local importa, es necesaria una evaluación con enfoque local, basada idealmente en los efectos directos del mercado laboral local en los participantes y en un grupo de comparación. Algo de importancia primordial es considerar si las economías de escala hacen que sea demasiado costosa la evaluación de programas en algunas localidades. El estudio de Leitner sobre el Pacto Territorial para el Empleo en Viena demostró que es factible técnicamente una evaluación rigurosa utilizando un diseño cuasi-experimental basado en un procedimiento de emparejamiento* de la propensión a participar** (*propensity score matching*) para construir grupos de comparación de participantes y de no participantes. Mosley también observó otras precondiciones para emprender dicha evaluación con éxito: la evaluación es una actividad obligatoria para el Fondo Social Europeo, el costo no era un obstáculo en el caso de Viena, existía la pericia necesaria para evaluar, se tenía fácil acceso a los microdatos de los registros de la seguridad social y se autorizó el tiempo suficiente para efectuar una evaluación adecuada.

* N. del T. Se dice de dos grupos (a, b) entre cuyas observaciones se ha establecido una correspondencia biunívoca a priori, formándose pares $(x_{a1}, x_{b1}) (x_{a2}, x_{b2}) \dots (x_{an}, x_{bn})$ de observaciones. El emparejamiento no necesariamente implica la existencia de correlación.

** *Score matching* es la probabilidad estimada de participar en el programa.

Evaluación de las microfinanzas

El artículo de investigación de Gary Woller —*Revisión de las metodologías para la evaluación del impacto en programas para el desarrollo de la microempresa*— brinda una amplia revisión de las evaluaciones de planes para microempresas tanto en las economías de la OCDE como en economías menos desarrolladas (EMD). Se tomaron en consideración 88 estudios de evaluación en total, consistentes en 67 evaluaciones de impacto independiente en 31 EMD, y 20 evaluaciones en 2 países de la OCDE, 19 de las cuales fueron en los Estados Unidos.

Woller observa que los planes de microfinanzas tienen impactos que pueden afectar al individuo, a la empresa, a las unidades domésticas y a la comunidad. Un espectro de factores económicos, sociales, de las políticas y específicos de la ubicación chocan con esos impactos. El artículo proporciona una tipología de los enfoques de evaluación que se han aplicado para evaluar las microfinanzas, desde experimentos aleatorios mediante un continuum de rigor decreciente, hasta investigación de mercado dirigida a guiar las decisiones de la gerencia. A su vez, se revisan los méritos y deméritos de cada enfoque. Vale la pena destacar que de 20 evaluaciones de microfinanzas en economías de la OCDE, sólo en siete se utilizó un grupo de control, en comparación con 73 de 90 evaluaciones de programas operados en EMD. Woller observa que aunque la falta de grupos de control excluye la inferencia causal de los efectos directos del programa, las evaluaciones sin grupos de control pueden aun así contribuir a lograr determinadas funciones administrativas. Por ejemplo, pueden dar seguimiento al progreso del usufructuario, valorar los efectos directos relativos entre diferentes segmentos del mercado y calcular los costos unitarios de resultados o productos específicos del programa.

Otra cuestión interesante que plantea este artículo está relacionada con la colocación no aleatoria de los programas. Los gestores del programa pueden ubicar un plan basándose en consideraciones que podrían tener influencia independiente sobre el desempeño del programa. Por ejemplo, unas buenas condiciones logísticas, que afectan el hecho de que los administradores puedan visitar o no un plan sin dificultad, podrían influir en la capacidad de las microempresas para comercializar sus productos. Esto puede llevar a resultados del programa diferentes a los que hubiera logrado si el mismo programa hubiese estado situado en otra parte. Para contrarrestar este problema, la ubicación del programa se puede asignar al azar. Dado que es probable que las ubicaciones de acogida potenciales de un programa que excedan sus capacidades, no necesariamente se carece de carácter práctico al elegir al azar donde se va a establecer el siguiente plan. Woller cita el estudio de una evaluación en el norte de Tailandia que demostró que no realizar la ubicación no aleatoria del programa (ni la autoselección del participante) condujo a una significativa sobreestimación del impacto.

El carácter fungible del uso que se hace de los préstamos se muestra como un problema particular en la evaluación de los planes de microfinanciamiento. Por ejemplo, los fondos del préstamo se podrían utilizar para propósitos distintos al desarrollo de la empresa, mientras que los reembolsos del préstamo podrían no provenir del flujo de caja de la empresa. El carácter fungible del préstamo agrava el problema de trazar una conexión directa entre la recepción de un préstamo y los cambios en el ingreso, el consumo, la acumulación de activos, etcétera, del prestatario/unidad doméstica. El artículo indica que para abordar este problema, las preguntas de la encuesta podrían tratar de obtener información sobre los usos de los fondos prestados y las fuentes para el pago del préstamo.

Woller hace importantes observaciones sobre lo engañosa que puede ser la presentación de los hallazgos de la evaluación. Por ejemplo, la tasa de reembolso del préstamo es un indicador de la calidad de la cartera de uso común. No obstante, esta medición puede omitir información importante contenida en un indicador superior, como la parte de la cartera en riesgo. Las valoraciones también suelen descuidar el reporte de las cifras de beneficiarios que abandonan el programa. Sin embargo, el porcentaje de los que abandonan el programa puede ser un contundente indicador de si los participantes consideran que el programa vale la pena o no. Hay también varias maneras de reportar mal los ingresos de la empresa. Por ejemplo, muchas veces se describe el nivel de ingreso de la empresa en lugar de los cambios en el ingreso; asimismo, a veces se reporta el ingreso sin hacer referencia a cómo se relaciona este ingreso con el tiempo trabajado y el capital invertido. También podría quedar confuso qué abarca en realidad el ingreso que se ha medido. Como respuesta a estos problemas, y como un principio general para el encargo de evaluaciones, se debería exigir que los evaluadores revelaran por completo las metodologías aplicadas. Habría que estipular que se hagan explícitos los inconvenientes metodológicos de los enfoques elegidos, las dificultades que se han encontrado, la lógica de la elección de los indicadores del desempeño, las áreas de valoración subjetiva y las posibles interpretaciones conflictivas. Como insiste Woller, poner de manifiesto todo lo anterior es crucial para que los diseñadores de políticas públicas y otras personas que no tienen por qué ser especialistas en evaluación puedan hacer una interpretación informada de los hallazgos de la evaluación.

Panel de debate

El artículo de Alice Nakamura relata las discusiones que tuvieron lugar en el debate final de la Conferencia, que reunió a seis eminentes panelistas (el profesor Edward Hill de la Universidad del Estado de Cleveland, el honorable Henry B. McLeish, ex primer ministro de Escocia, el doctor Stephen Wandner, del Departamento del Trabajo de los Estados Unidos, el profesor Philip Davies, del Consejo de Ministros del Reino Unido, y la profesora Alice Nakamura, de la

Universidad de Alberta). La discusión se centró en el compromiso del gobierno con la evaluación y en cómo mejorar la situación actual. Las cuestiones que se trataron fueron desde la elección de indicadores del desarrollo económico hasta las tensiones entre la tendencia de los diseñadores de políticas a reservarse los detalles de los procesos de evaluación, los vínculos entre el diseño de políticas y la academia, el libre acceso de ciertas prácticas de evaluación en los Estados Unidos, los incentivos posibles para alentar la práctica de una buena evaluación, la integración de la recolección de datos, la práctica metodológica, el diseño y la implementación de las políticas públicas, y las maneras de mejorar la transmisión de información sobre la evaluación a los medios de comunicación.

Temas clave que surgen de la conferencia

Diseño estratégico de la evaluación

Queda claro por las ponencias y la riqueza de las prácticas que describen, que pocos gobiernos hacen uso de la heterogeneidad implícita en las distintas formas regionales/locales de diseño e implementación de programas. La variación en la implementación y el diseño de las políticas públicas a través de las regiones y las áreas locales, así como la ubicación aleatoria de los programas, brinda la oportunidad de verificar la eficacia de diferentes diseños de políticas. Es obvio que se necesita trabajar más en la incorporación de la evaluación a la formulación de las políticas de manera estratégica. Incluso en el plano de cada programa hay amplias oportunidades de generar abundante información sobre las políticas a través de las variantes locales de modalidades de implementación. Por ejemplo, una autoridad regional que patrocina un programa de microcréditos puede variar características clave del diseño del programa —tales como los planes de reembolso de préstamos y las técnicas para recibir préstamos en grupo o individualmente— a través de una serie de áreas locales. De manera igualmente estratégica, pero con miras a mejorar el acceso a datos, el artículo de Bondonio indica que la Unión Europea podría reasignar parte de los programas de asistencia a áreas locales que coinciden exactamente con zonas sobre las que se cuenta con la información estadística detallada.

La supuesta intangibilidad de importantes efectos directos en el desarrollo local

Durante la conferencia de Viena se expresó reiteradamente la idea de que muchos efectos directos del desarrollo local son intangibles o difíciles de cuantificar. Es más, esta aseveración es frecuente entre los profesionales del desarrollo local. Por ejemplo, a veces se sostiene que las agencias del tercer sector son capaces de crear una gama de efectos directos benéficos que pueden ser

difíciles de calibrar. Se dice que dichos beneficios incluyen un incremento del capital social, mayor compromiso de la comunidad con iniciativas de regeneración, confianza en sí mismos de los participantes y la generación de información local (superior). Pero puede ser que se haya exagerado la pretendida intangibilidad de los efectos directos del desarrollo local. En primer lugar, algunos efectos directos pueden ser difíciles de medir, sobre todo porque todavía no se llega a un consenso sobre las bases conceptuales. Tal es el caso del polémico concepto de “capital social”. El reto de la evaluación en este caso podría estar relacionado en primer lugar con si los patrocinadores del programa tenían un concepto significativo del capital social, es decir, si la lógica del programa era coherente. Los programas que pretenden acrecentar la “cohesión de la comunidad” pueden no ser susceptibles de evaluación, no por el hecho de que algunos componentes de la cohesión no se pueden medir, sino porque los autores de los programas no han especificado desde el inicio a qué se refieren con cohesión de la comunidad. Otros tipos de beneficios menos tangibles en realidad podrían ser mensurables. Por ejemplo, la confianza en sí mismas y las opiniones locales de las iniciativas de desarrollo se pueden dimensionar a través de encuestas. Cuando se puede formular específicamente un efecto directo, suele ser mensurable de alguna manera, directamente o por medio de sustitutos.

Tensiones entre el enfoque “científico” y otros enfoques de evaluación del desarrollo local

Varios de los artículos y algunas de las intervenciones en la conferencia indican un falso antagonismo en algunos sectores entre los enfoques de evaluación econométricos y los participativos. Ambos deberían representar puntos en el *continuum* de una evaluación, y ambos proporcionar percepciones de diferentes tipos. La ponencia de Woller describe las críticas a las técnicas científicas para evaluar el microfinanciamiento. Se considera que estas críticas contienen afirmaciones de que dichos métodos: no logran capturar toda la complejidad de las relaciones causales; intentan cuantificar lo que no se puede cuantificar; potencian a los tecnócratas, y no consiguen tener por resultado la acción de o para los grupos pobres. No todos estos argumentos son convincentes. Por ejemplo, es difícil entender porqué el uso de métodos científicos tiene más probabilidades de implicar un concepto erróneo de la causalidad que otras formas de evaluación. Es por lo menos tan probable que los analistas letrados en estadística sean conscientes de las interacciones causales complejas como los analistas versados en otras formas de indagación. No obstante, las críticas en contra de los denominados métodos científicos sí indican a los investigadores la necesidad de estar atentos al contexto más amplio en el que se está evaluando un programa y de prestar la debida consideración a la evidencia procedente de los que participan en el programa.

El desplazamiento y la necesidad de evaluar las políticas más pertinentes

El desplazamiento fue un tema muy discutido en la conferencia. El desplazamiento adopta varias formas y está presente en diversos aspectos de las políticas de desarrollo local. Por ejemplo, existe la preocupación de que combatir la delincuencia en una localidad podría llevar a que ésta se desplazara a otra; y un proyecto de asociación podría desplazar recursos otras actividades como el tiempo de los individuos participantes.

Hay graves problemas metodológicos para medir la magnitud del desplazamiento. Para poner el caso del desplazamiento empresarial: una complicación particular consiste en que la medición del desplazamiento suele depender de encuestas a las empresas que pueden haber causado desplazamiento en otras. A veces, las empresas desplazadas ya no están en funcionamiento y las perspectivas de los dos grupos de empresas podrían diferir sistemáticamente. Además, no hay sólo un indicador estándar del desplazamiento, sino una serie de condiciones específicas del contexto que son las que determinan la gravedad del desplazamiento. Por ejemplo, es probable que el desplazamiento sea mayor en el tipo de empresas que predominan en las localidades pobres, porque se trata de firmas en sectores maduros de bajo crecimiento en las que los requerimientos de habilidades y de capital son limitados, y en las que las barreras de entrada son bajas. Además, la situación de la demanda y la oferta locales son cruciales para medir la magnitud de los efectos de desplazamiento. Y aunque es probable que es el desplazamiento sea más agudo en el plano local, no se limita necesariamente a los confines de una localidad en particular: empresas en otros lugares también pueden ser desplazadas. Es también consustancial al desplazamiento que haya diferencias considerables en la geografía de los mercados atendidos por empresas de diversos sectores (es más probable que las del sector servicios tengan una orientación local, en especial, las que prestan servicios personales). El desplazamiento también variará dependiendo de si los productos de las empresas en el mismo sector son sustitutos cercanos o no. A la luz de estas observaciones, no es sorprendente que las estimaciones del desplazamiento hayan variado significativamente [hasta en un cien por ciento en algunos planes de apoyo empresarial (OCDE, 2003)].

En resumen, es sabido que la complejidad del fenómeno aconseja tratar con cautela las estimaciones del desplazamiento. También es sabido bajo que condiciones estandarizadas es más probable asociar los niveles altos o bajos de desplazamiento. Por consiguiente, lo más necesario tal vez sea una evaluación ulterior relacionada con las políticas para establecer medidas que podrían mitigar el desplazamiento. Por ejemplo, los programas para ayudar a las empresas a vender en mercados fuera de la zona y a elevar el monto promedio de la inversión en nuevas empresas podrían, en conjunto, reducir el desplazamiento (OCDE, 2003). Dicha investigación podría rendir al menos tanta información pertinente

a la política que los intentos cada vez más detallados de mantener bajo control la escala y amplitud del desplazamiento, que son de todos modos elusivos y altamente específicos del contexto.

Recomendaciones para las políticas públicas: fundamentos

En la Segunda Conferencia Ministerial de la OCDE sobre las pequeñas y medianas empresas (PYMEs) en junio de 2004, se reconoció que la evaluación de las políticas era clave para apoyar a las PYMEs innovadoras. Un ponente de la conferencia hizo la siguiente observación, que en realidad se puede aplicar a toda la gama de las políticas públicas:

“Los diseñadores de políticas deberían ser capaces de responder cuatro preguntas sobre sus políticas. ¿Son claras y coherentes? ¿Tienen objetivos claros? ¿Cuáles son las metas? ¿Se evalúan? Si no se evalúan, se están engañando a ustedes mismos diciendo que sí logran sus objetivos”. (David Storey, *CORDIS News*, 2004-06-07)

En términos prácticos y con base en los consejos de expertos presentados en este volumen, la implementación de las siguientes recomendaciones contribuirá a una cultura de la evaluación que producirá evaluaciones de mejor calidad y más útiles para tomar decisiones informadas sobre desarrollo local.

Para el gobierno central y local

Modernizar y racionalizar el desarrollo y la implementación de programas y políticas públicas

La tarea de la evaluación es enormemente complicada cuando una serie de programas gubernamentales en curso se superponen o en parte coinciden. Cuando los programas se reemplazan previamente a la evaluación o se evalúan antes de que sea evidente todo su impacto, se pueden despilfarrar recursos públicos. Y cuando las iniciativas son demasiado numerosas, invertir en el aprendizaje de las políticas puede parecer improductivo.

Hacer explícito, al nivel más alto posible, el compromiso con la evaluación y su importancia. Asimismo, debe haber un compromiso con la difusión pública de los hallazgos de la evaluación

Es vital el franco reconocimiento de la importancia de la evaluación por parte de diseñadores de políticas y líderes de agencias gubernamentales versados en la materia. Es más probable que se disponga de recursos humanos y financieros para la evaluación una vez que se ha hecho evidente dicho reconocimiento. Desde la concepción del programa debería reservarse una proporción de los presupuestos

totales con fines de evaluación. Por ejemplo, en el Reino Unido, se ha separado una parte de los recursos disponibles para ciertos programas de regeneración con fines de evaluación (DETR, 2001); y para todas las propuestas de políticas públicas en Australia se requiere la inclusión de una estrategia de evaluación.

Para alentar una mayor apertura hacia la evaluación, debe quedar claramente establecido que la intención de la evaluación es mejorar la calidad de las políticas públicas. Los reportes gubernamentales regulares del desempeño deberían emanar ampliamente de los resultados de la evaluación. Las evaluaciones que revelan problemas en un programa determinado no se deberían ver como bases automáticas para poner fin a un programa, sino como una herramienta que proporciona la base para mejorar esa política en concreto. Además, la divulgación de los estudios de evaluación y sus resultados, posiblemente de manera anónima, puede servir como una forma de control de calidad de la evaluación y como un medio para estimular el análisis académico pertinente a las políticas públicas.

Los gobiernos también deberían comprometerse a hacer del dominio público los datos brutos de la evaluación de manera que sean asequibles y accesibles a los usuarios potenciales. Por ejemplo, como parte de las evaluaciones de un gran número de programas del mercado laboral patrocinados por el Departamento del Trabajo de los Estados Unidos se han puesto a disposición en línea datos fundamentales a un bajo costo.

Asimismo, los gobiernos podrían crear o apoyar a instituciones que proporcionan capacitación y asistencia técnica a la evaluación, especialmente a autoridades subnacionales, o que sirven como centros de excelencia en evaluación. Los gobiernos también podrían establecer o apoyar foros (conferencias, seminarios, talleres y mesas de trabajo, sitios en Internet, etcétera) que promuevan el debate y la difusión de información sobre resultados y métodos de evaluación. Una importante función de los foros de evaluación o de los centros de excelencia podría ser fomentar el entendimiento entre la investigación y las comunidades objeto de las políticas públicas.

Decretar evaluaciones cuando se proporciona financiamiento público

Una razón fundamental por la cual varios programas patrocinados estatal y federalmente han sido evaluados detalladamente en los Estados Unidos, como el programa de Asociaciones para la Ampliación de la Industria Manufacturera, es que es obligatoria cuando se hace uso de los recursos públicos federales. Los requisitos de evaluación del gobierno central (o regional) también podrían implicar la recolección estandarizada de datos por las agencias subnacionales (o insertar protocolos para la recolección estandarizada de datos en el diseño del programa). Esto podría ayudar a comparar los impactos a lo largo de los diferentes escenarios locales de algunos de los programas más utilizados.

Aclarar los resultados que se van a evaluar durante el diseño del programa y sin duda antes de la implementación del programa

A los gestores del programa en perspectiva se les debería pedir que detallen los pasos requeridos para evaluar el impacto del programa, incluyendo la recolección de datos con fines de evaluación. La preparación de un borrador del proceso de evaluación tendría que involucrar también idealmente a la agencia encargada de realizar la evaluación. Entre los aspectos que se debería aclarar están la elección de metodologías alternativas y los intercambios que implican, los tipos de impactos que se estudiarán, los tipos de datos que se recolectarán, los indicadores que se usarán y cómo se reportarán los resultados. También es importante decidir cuándo se llevarán a cabo las evaluaciones ya que los diferentes programas tienen periodos de gestación distintos. Prepararse para emprender una evaluación de calidad es en sí un ejercicio pedagógico que posee el potencial de mejorar cómo se concibe el diseño del programa. Por ejemplo, las iniciativas de desarrollo local suelen implicar objetivos múltiples (sociales, ambientales, económicos, etcétera), pero el logro simultáneo de objetivos múltiples puede ser irreal. La especificación previa de los objetivos podría contribuir a evitar inversiones en programas inadecuados.

Evaluar cuando es probable que se obtengan beneficios netos por hacerlo

Algunos tipos de programas han sido evaluados de manera rigurosa en múltiples ocasiones en diversos países. Nuevas evaluaciones de programas similares tal vez no representarían una buena inversión pública. En este volumen, Smith señala que el “tiempo empleado en la lectura de bibliografía sobre buenas evaluaciones de programas similares puede redituarse resultados más útiles que una evaluación débil realizada con base en datos escasos por un evaluador poco calificado utilizando métodos inapropiados”. Obviamente, también tiene poco sentido invertir recursos en la evaluación de programas que de todas maneras serán descontinuados. Además, en el caso de programas pequeños, los beneficios potenciales obtenidos podrían no justificar el costo de evaluaciones sofisticadas. Asimismo, la evaluación podría ser superflua si los datos existentes son de baja calidad o si los tamaños de las muestras son demasiado pequeños como para generar resultados estadísticamente válidos.

Elegir la técnica de evaluación a la luz del tamaño y la naturaleza del programa en cuestión

Los estudios de programas importantes deberían utilizar diversos métodos: muestreo aleatorio, evaluación cuasi-experimental, entrevistas con los beneficiarios, enfoques participativos que involucren a todas las partes interesadas. Las evaluaciones de los resultados que utilizan enfoques experimentales son

costosas, además de ser intensivas en datos y requerir de una pericia estadística particular. Como tales, estas técnicas no suelen ser adecuadas para los niveles inferiores gubernamentales que operan programas relativamente pequeños. Cuando las aplican gobiernos estatales o municipales (subnacionales), estas técnicas se deberían emplear para evaluar programas más grandes (o programas piloto que podrían ampliarse significativamente), en los que los beneficios potenciales de las mejoras de la calidad del programa son mayores.

Insistir en la publicación de los reportes de evaluación en su totalidad

Deben describirse cabalmente la elección de los métodos y parámetros de evaluación utilizados, los inconvenientes metodológicos y las áreas de apreciación subjetiva. Los estudios de evaluación deberían ser claros sobre lo que puede y no puede cuantificarse. Dado que los resultados pueden expresarse de diferentes maneras utilizando diferentes variables, se debería declarar explícitamente cómo se ha optado por expresar los resultados.

Ejecutar la evaluación utilizando tanto expertos independientes como personal interno

Si personal interno administra los sondeos de los beneficiarios del programa, en particular de los usuarios con los que se ha trabajado directamente, existe un elevado riesgo de sesgo en las respuestas. Los programas deberían ser evaluados también por expertos externos independientes, posiblemente de una oficina de auditoría, o en colaboración con ellos. Idealmente, el organismo que efectúa la evaluación trabajaría con los operadores del programa, pero la continuidad de los contratos no dependería del que patrocina el programa. No obstante, en este caso la cuestión es la independencia del personal interno que participa en las evaluaciones. El personal necesita estar a salvo de cualquier forma de sanción en los casos en que los hallazgos de la evaluación resulten ser negativos.

Alentar a los evaluadores a considerar las implicaciones de los hallazgos de la evaluación para diversos cursos posibles de las políticas

Por ejemplo, una evaluación con una perspectiva orientada al futuro podría destacar el hecho de que la disminución del rendimiento podría aumentar si un determinado programa se expande.

Invertir en la recolección de datos adecuados de la localidad

No suele haber disponibilidad de datos administrativos sobre áreas geográficas pequeñas. Y, a veces, las diferentes áreas del gobierno recolectan datos de unidades

geográficas a diferentes escalas. Es necesario contar con sistemas estadísticos que provean datos de fácil acceso sobre unidades geográficas pequeñas que permanecen estables en el tiempo. Es esencial recolectar bases de datos que sean pertinentes a los objetivos de las políticas, que se puedan obtener de localidades y empresas objetivo y de comparación, y a los que también se les pueda dar seguimiento a lo largo del tiempo. Recolectar los datos en el punto de entrega del programa puede ser mucho menos costoso que recolectar los datos *ex post*. Especialmente en las fases de prueba piloto del desarrollo del programa, la asistencia al programa se podría dirigir a áreas para las que se disponga de información estadística detallada. Algunos países están empezando a abordar el problema de los datos en las áreas locales. Por ejemplo, como lo describe Robson en este volumen, el gobierno del Reino Unido ha invertido en el establecimiento de una base estadística de datos nacional por distritos (neighbourhood.statistics.gov.uk), valiéndose de las autoridades locales por distrito electoral como unidad de reporte.

Para los gobiernos centrales

Ser estratégico en el diseño de los apoyos al desarrollo económico local con el fin de facilitar la identificación de las localidades del grupo de control

A veces, las localidades que albergan las iniciativas de políticas en particular tienen pocas localidades de comparación, sobre todo, si la política en cuestión es altamente incluyente. Por ejemplo, si los programas se implementan en todas las localidades que tienen una tasa de disminución del desempleo, los únicos lugares que quedan para hacer la comparación serán aquellos que probablemente tengan características disímiles importantes.

Apuntar a conseguir un compromiso coherente entre los diferentes esferas del gobierno

Es poco frecuente que diferentes esferas de gobierno se impliquen conjuntamente en el financiamiento, diseño e implementación de las evaluaciones. Por ejemplo, en los Estados Unidos han sido muy raras las evaluaciones intergubernamentales genuinas (Morra, 1997). El gobierno federal de los Estados Unidos ha tenido el liderazgo, aunque los estados se han vuelto cada vez más activos en la evaluación. En Francia, para evaluar los contratos de planeación entre el gobierno central y las regiones, cada región desarrolló sus propios métodos y procedimientos. La necesidad de coordinar las evaluaciones se articuló en una circular del Primer Ministro en 2000 (Heddebaut, 2000).

Considerar el proporcionar financiamiento a las evaluaciones elaboradas localmente

Este capítulo y otros más de este libro describen por qué es probable que los organismos locales subinviertan en evaluaciones de alta calidad. Las exigencias de que algunos tipos de programas logren la autonomía financiera también podría desalentar el gasto en evaluación. En consecuencia, hay razones para que los gobiernos centrales proporcionen apoyo financiero o de otra índole a las evaluaciones realizadas localmente. Dicho apoyo podría incluir también la capacitación de evaluadores.

Desarrollar estándares y lineamientos claros para la evaluación

Las autoridades públicas deberían tratar de desarrollar estándares y lineamientos de evaluación. Las autoridades centrales de algunos países ya lo hacen (por ejemplo, la Oficina de Contabilidad General de los Estados Unidos). Entre otros beneficios, los estándares y lineamientos de evaluación podrían ayudar a los gobiernos locales a establecer las normas metodológicas que se esperan de las compañías subcontratadas para realizar evaluaciones. En realidad, el apego a los estándares de evaluación podría convertirse en una condición para el apoyo y los contratos del programa. Además de mejorar la calidad de la evaluación, una mayor uniformidad en la práctica de la evaluación también podría contribuir a remediar la situación actual en la que los resultados de los programas no suelen ser comparables debido a la utilización de métodos de evaluación opacos o no estandarizados. Los estándares indicados podrían aplicarse incluso al diseño de las encuestas.

Para los gobiernos locales

Las autoridades locales deberían desempeñar un papel activo en el desarrollo de estudios de caso y en las encuestas a los participantes del programa

Es probable que las encuestas sean más confiables cuando abordan programas que proporcionan servicios reales en lugar de apoyo financiero. Las encuestas pueden ser valiosas para ayudar a esclarecer porqué los programas han funcionado o no y qué debería hacerse para mejorar la eficacia mediante diseños y modos de implementación alternativos. Cuando se realiza la encuesta, es necesario plantear las preguntas de manera sumamente específica. Las preguntas deben abordar el contrafactual y centrarse en las características clave del diseño del programa. Las encuestas se deben basar en muestras suficientemente grandes para poder extraer de ellas inferencias válidas.

Buscar la asociación con organismos académicos

La evaluación es una disciplina técnica y especializada. A menudo, las autoridades locales no poseen la necesaria pericia interna. Para realizar evaluaciones rigurosas es crucial tener conocimientos de economía, teoría estadística y características de los datos. Una manera de salvar la brecha de las capacidades es formar asociaciones de evaluación con instituciones académicas, o en general, recurrir a sus habilidades y pericia. Los académicos también pueden estar dispuestos a analizar datos a bajo costo si es probable que los resultados sean de interés para la comunidad académica. Asimismo, debe impulsarse también la publicación de las investigaciones pertinentes. Cuando los gobiernos estatales y municipales subcontratan estudios sobre evaluación con empresas privadas de consultoría puede ser importante recurrir a expertos académicos para garantizar términos de referencia adecuados. La pericia académica también puede ayudar a comparar la calidad relativa de diferentes estudios de evaluación, dado que las conclusiones de las políticas podrían estar basadas en trabajos endebles desde una perspectiva metodológica, teórica o de la calidad de los datos. Las autoridades locales también podrían crear asociaciones de evaluación a través de asociaciones nacionales de gobiernos locales o con ellas. Dichas asociaciones intergubernamentales podrían proponerse la divulgación de los hallazgos de la evaluación a organismos estatales y municipales y garantizar una asesoría técnica competente.

Notas

1. Los diseñadores de políticas tienen que escoger entre asignar recursos públicos a usos diferentes. En un modelo de libro de texto, los políticos locales a nivel mundial asignan recursos a diferentes programas basándose en el conocimiento del costo *marginal* del logro de objetivos comunes determinados para diferentes tipos de programas. Por ejemplo, si la creación de empleos fuera la meta principal de la política de desarrollo local, se podría lograr una asignación de recursos económicamente eficiente si los diseñadores de políticas tuvieran información sobre los costos *marginales* de la creación de un puesto de trabajo mediante programas tipo atracción de inversiones, subvenciones para la puesta en marcha de empresas, capacitación, etcétera. Se logra una asignación eficiente cuando los costos marginales de la creación de empleos son los mismos en todos los programas (Storey, 1990). No obstante, en la práctica, la mayor parte de las evaluaciones brindan información sobre el costo *promedio* de la creación de empleos. Además, el costo marginal de la creación de empleos variará en el transcurso del tiempo dependiendo, por ejemplo, de la escala y la duración de un programa y de los rasgos clave del mercado de trabajo local. El ideal de libro de texto, pues, necesitaría un ciclo constante de evaluación compleja de muchos tipos de programas y de hecho, es inalcanzable.
2. El asociacionismo se trata a menudo como un bien por sí mismo. La discusión de posibles costos de oportunidad o efectos de desplazamiento (véase nota 4) derivados del trabajo en asociación, pareciera ser escasa.

3. Bartik señala también que no está claro si la rentabilidad de etiquetar recursos para el desarrollo de áreas es más alta en zonas con mayor o menor marginación. No obstante, es poco probable que haya una sola respuesta a esta pregunta. En la práctica, el que un programa en particular llegue a tener mayor o menor impacto en áreas más o menos marginadas, depende del tipo de programa. Están por supuesto en uso muchos tipos diferentes de programas para el desarrollo de zonas. Algunos tipos de programas podrían ser más efectivos de acuerdo con algunos criterios de evaluación que otros, y las variaciones en este sentido pueden ser sensibles al grado de marginación encontrado. Por ejemplo, en los planes de apoyo empresarial en zonas con mayores niveles de bienestar frecuentemente se atenderá a las empresas más grandes y a los empresarios mejor calificados. Esto puede asociarse con una mayor longevidad de la empresa y con un mayor efecto multiplicador. El desplazamiento empresarial también puede ser menor en las zonas más ricas que en las pobres. Por lo que el apoyo empresarial puede funcionar bien, según estos parámetros de evaluación, en localidades acaudaladas. No obstante, como es más probable que los empresarios mejor calificados hayan emprendido negocios ante la falta de apoyos empresariales, el “peso muerto” asociado con dichos planes podría ser más alto en las localidades con mayor riqueza (OCDE, 2003).
4. El peso muerto se refiere al grado en que habría habido resultados sin intervención de las políticas. Por ejemplo, a falta de una iniciativa de asociación formal, otras instituciones o individuos habrían podido unirse de cualquier otra forma para alcanzar objetivos similares. El desplazamiento se refiere esencialmente a la pérdida de producción, empleos u oportunidades derivada de un programa o una política. El desplazamiento toma muchas formas a lo largo de la gama de mercados. Por ejemplo, las intervenciones pueden hacer uso de recursos —como herramientas o terrenos— que se vuelven inaccesibles para usos alternativos. O el precio de estos recursos puede subir. Y desde una perspectiva de desarrollo comunitario, los programas para fomentar la formación de redes locales podrían conducir posiblemente a una reducción en la actividad de formación de redes a través de una zona más amplia [Amstrong et al. (2002)]. Las intervenciones financiadas mediante impuestos generales tendrá también un efecto de desplazamiento en el consumo privado o en la inversión en otras partes de la economía (aunque estos efectos pueden ser contrarrestados con mejoras del lado de la oferta en la economía). Si el desplazamiento sucede entre localidades en lugar de al interior de la localidad, entonces su importancia desde la perspectiva del desarrollo de zonas también dependerá de si el desplazamiento ocurre en áreas que son más o menos acomodadas que la localidad objetivo. Los efectos de sustitución son descritos frecuentemente en conexión con la política del mercado laboral. Por ejemplo, la sustitución se presenta cuando las personas que reciben capacitación ocupan las vacantes que de otra manera hubieran sido ocupadas por personas no capacitadas.
5. Sin embargo, es frecuente que los lineamientos recomienden a los evaluadores acciones que son casi inoperables. Por ejemplo, el Erario de Su Majestad recomienda que “Los estudios de evaluación deberían proporcionar información siempre sobre el efecto del desplazamiento local y nacional (UK)”. Esto es exigir demasiado.
6. Malan (2002), en relación con esto, reporta sobre una importante evaluación *ex post* de los programas de la Comisión Europea Objetivo 2, 1989-1993. Señala que tanto el Fondo Europeo para el Desarrollo Regional como el Fondo Social Europeo estaban ansiosos por estimar las sinergias que se daban entre programas co-ubicados. Sólo tres de los 25 evaluadores involucrados intentaron aplicar la metodología de evaluación de la Comisión Europea (MEANS) con este fin, argumentando que la tarea era muy compleja y que consumía demasiado tiempo.
7. Por ejemplo, Barkley (2003) documenta que los diseñadores de políticas en el ámbito estatal de los Estados Unidos no se han tomado el tiempo necesario para juzgar la efectividad de programas innovadores de participación en el capital social en zonas

rurales. Dichos programas pueden involucrar fondos de inversión con ciclos de negocio de cinco años o más. Muchas veces los diseñadores de políticas han introducido nuevas generaciones de planes antes de obtener una evaluación completa de los programas existentes.

8. El término “empresa social” usualmente se refiere a empresas que se proponen lograr objetivos sociales usando recursos de diversas fuentes públicas y privadas. La combinación de financiamiento público y de otras fuentes con ingresos derivados de transacciones mercantiles es el rasgo determinante que distingue a las empresas sociales de las organizaciones tradicionales sin fines de lucro.

Referencias bibliográficas

- ARMSTRONG, H.W., B. KEHRER, P. WELLS y A.M. WOOD (2002), “The Evaluation of Community Economic Development Initiatives”, *Urban Studies*, vol. 39, núm. 3, pp. 457-481.
- BARKLEY, D.L. (2003), “Policy Options for Equity Financing for Rural Entrepreneurs”, en *Main Streets of Tomorrow: Growing and Financing Rural Entrepreneurs*. Center for the Study of Rural America, Federal Reserve Bank of Kansas City, octubre 2003, pp. 107-126.
- BARTIK, T.J. y R.D. BINGHAM (1997), “Can Economic Development Programs be Evaluated?” en R.D. Bingham, y R. Mier (comps.), *Dilemmas of Urban Economic Development*, Sage Publications.
- BOARNET, M.G. (2001), “Enterprise Zones and Job Creation: Linking Evaluation and Practice”, *Economic Development Quarterly*, vol. 15, núm. 3, agosto 2001, pp. 242-254.
- BONDONIO, D. y J. ENGBERG (2000), “Enterprise Zones and Local Employment: Evidence from the States’ Programms”, *Regional Science and Urban Economics*, 30, pp. 519-549.
- CANADÁ (2000), *Study of the Evaluation Function in the Federal Government*, Treasury Board of Canada Secretariat, Ottawa.
- COWLING, M. y R. HAYWARD (2000), *Out of Unemployment*, Research Centre for Industrial Strategy, Reino Unido.
- DETR, DEPARTMENT OF THE ENVIRONMENT, TRANSPORT AND THE REGIONS (2001), *A Review of the Evidence Base for Regeneration Policy and Practice*, Londres.
- EISINGER, P. (1995), “State Economic Development in the 1990s: Politics and Policy Learning”, *Economic Development Quarterly*, vol. 9, pp. 146-158.
- FOLEY, P. (1992), “Local Economic Policy and Job Creation: A Review of Evaluation Studies”, *Urban Studies*, vol. 29, núm. 3 y 4, pp. 557-598.
- HEDDEBAUT, O. (2000), “Decentralization and Evaluation: the Problems Caused by the Various Levels of Decision and the Length of the Observation Period”, ponencia presentada en la Cuarta Conferencia de la Sociedad Europea de Evaluación, Lausanne, octubre 12-14.
- HM TREASURY (1995), *A Framework for the Evaluation of Regeneration Projects and Programmes*, Londres.

- MALAN, J. (2002), "Translating Theory into Practice: Lessons from the ex post Evaluation of the 1989-1993 Objective 2 Programmes", ponencia presentada en la Conferencia de 2002 de la Sociedad Europea de Evaluación, Sevilla, España, octubre 10-12.
- MEIER, J.K. y J. GILL (2000), *What Works: A New Approach to Program and Policy Analysis*, Westview Press, Oxford.
- METCALF, H., T.V. CROWLEY, T. ANDERSON y C. BAINTON (2000), *From Unemployment to Self-Employment: The Role of Micro-Finance*, International Labour Office, Londres.
- MORRA, L.G. (1997), "Evaluation in the United States: Cooperative but not Intergovernmental", en Rieper y Toulemonde: *Politics and Practices of Intergovernmental Evaluation*, Transaction Publishers.
- OCDE (1999), *Improving Evaluation Practices*, OCDE, París.
- OCDE (2003), *Entrepreneurship and Local Economic Development: Programme and Policy Recommendations*, OCDE, París.
- DEPARTMENT FOR TRANSPORT, LOCAL GOVERNMENT AND THE REGIONS (2002), "Lessons and Evaluation Evidence from Ten Single Regeneration Budget Case Studies", Reporte de mitad de periodo, Londres.
- RIEPER, O. y J. TOULEMONDE (1997), *Politics and Practices of Intergovernmental Evaluation*, Transaction Publishers.
- SCHMID, G., J. O'REILLY y K. SCHÖMANN (1996), "Theory and Methodology of Labour Market Policy and Evaluation: An Introduction", en Schmid, O'Reilly y Schömann (comps.), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar, Cheltenham.
- SHAPIRA, P. (2003), "Evaluating Manufacturing Extension Services in the United States: Experiences and Insights", en Shapira, P. y Kuhlmann, S., *Learning from Science and Technology Policy Evaluation*, Edward Elgar, Northampton, Massachusetts y Cheltenham, Reino Unido.
- STATE SERVICES COMMISSION (1999), "Looping the Loop: Evaluating Outcomes and Other Risky Feats", *Occasional Paper*, núm. 7, Wellington, Nueva Zelanda, junio.
- STOREY, D.J. (1990), "Evaluation of Policies and Measures to Create Local Employment", *Urban Studies*, vol. 27, núm. 5, pp. 669-684.
- THE AUDIT COMMISSION FOR LOCAL AUTHORITIES AND THE NATIONAL HEALTH SERVICE (1999), *A Life's Work: Local Authorities, Economic Development and Economic Regeneration*, septiembre, Reino Unido.
- WOOLCOCK, M.J.V. (1999), "Learning from Failures in Micro-finance", *The American Journal of Economics and Sociology*, 58 (1), pp. 17-42.

Capítulo 2

Aprendizaje de políticas mediante la evaluación: retos y oportunidades

Por
Ging Wong
Canadian Heritage y Universidad de Alberta

Contexto

Imagínese usted a un nuevo delegado asistiendo a su primera reunión del Comité de Desarrollo Económico y del Empleo a Nivel Local (LEED, por sus siglas en inglés) de la Organización para la Cooperación y el Desarrollo Económico (OCDE). Si bien está ansioso por compartir las experiencias de su país, también siente emoción ante la perspectiva de poder aprender de los otros participantes una variada gama de innovaciones económicas y sociales basadas en iniciativas locales. No se decepciona, ya que el Programa LEED le ofrece sólidos recursos analíticos y de información para discutir, grupos de trabajo para comparar experiencias y buscar respuestas a problemas de las políticas, y sobre todo, cientos de estudios de casos prácticos que documentan éxitos y fracasos importantes. Sin embargo, se siente frustrado ante tanta información. ¿Cómo destilar la esencia de esta riquísima fuente de información? ¿Cómo extraer las lecciones clave que podrían tener el potencial de mejorar la calidad de las políticas y operaciones públicas?

Ese delegado era yo, hace siete años. Desde entonces, junto a otros, he trabajado estrechamente con los directivos y la Secretaría del programa LEED en el desarrollo de evaluaciones rigurosas y sistemáticas de las prácticas actuales. En síntesis, buscamos promover una cultura de Evaluación, no como un fin en sí mismo, sino como una valiosa herramienta de información. La conferencia sobre la cual se basa este libro representa el lanzamiento de una iniciativa nueva y el comienzo de un nuevo diálogo.

Introducción

Permítame presentarle algunas ideas iniciales sobre el aprendizaje de políticas a través de la evaluación.

Si la calidad, la relevancia y la coyuntura son los distintivos de una buena evaluación, entonces nos será de utilidad el trabajo de la Sociedad Europea de Evaluación. Me sorprendió particularmente el tema de su conferencia anual del 2002: “Tres Tendencias en la Evaluación Contemporánea: Aprendizaje, Teoría y Evidencia”. Uno de los objetivos de la Conferencia de Viena fue explorar las

relaciones esenciales entre estos tres elementos y sus implicaciones en las políticas públicas de desarrollo económico y del empleo a nivel local. Por ello, concentraré mi análisis en el potencial de aprendizaje que se genera a partir de la evaluación.

Comenzaré citando el folleto de invitación a dicha conferencia:

“El aprendizaje es generalmente visto como un objetivo amplio de la evaluación. El tema del aprendizaje en el campo de evaluación empieza como una inquietud: los diseñadores de políticas debiesen aprender de ella. Pero en un mundo en donde la participación del Estado es limitada y frecuentemente en alianza con la sociedad civil, la evaluación como herramienta de aprendizaje no sólo pertenece al dominio de los diseñadores de políticas. Las comunidades, asociaciones y ciudadanos también utilizan la Evaluación para aprender: para comprender mejor las opciones disponibles y para ayudar a desarrollar nuevas formas de consensos inclusivos. Otro impulso que está reformulando el concepto de aprendizaje como una tendencia en el campo de la evaluación proviene de especialidades, tales como la administración del conocimiento, el aprendizaje organizacional y la memoria organizacional. Los evaluadores se preocupan cada vez más por no perder lo que ya se ha aprendido, por que no se pierdan los hallazgos de evaluaciones pasadas y que éstos sean accesibles y utilizables en el futuro”.

Esta cita describe elegantemente el poder de la información que genera la evaluación. Dicha información presenta evidencias de desempeño e impacto que deberán ser interpretadas en un contexto y marco teórico apropiado para las distintas audiencias objetivo que desean confirmar si las políticas e intervenciones están o no funcionando. También propone diversos objetivos en el campo de la evaluación, el porqué los gobiernos podrían interesarse por invertir en el desarrollo de sistemas de evaluación, a saber: para la rendición de cuentas públicas, promover procesos democráticos y establecer una base de conocimientos para la investigación y el desarrollo de políticas; todos estos son temas que deben recibir apoyo.

Sin embargo, el aprendizaje de políticas a través de la evaluación se topa con desafíos así como con oportunidades. A fin de explorarlos, este artículo comenzará examinando las oportunidades y usos actuales de la evaluación para luego llegar a las necesidades especiales de desarrollo y a los programas de empleo locales. Estos comentarios nacen de mis experiencias como antiguo director de la agencia de evaluación más grande del gobierno canadiense, así como de mi interacción con la comunidad internacional a través del Banco Mundial, la OCDE y los intercambios técnicos y políticas bilaterales entre Canadá y otros países.

Alcances y usos de las actividades de evaluación

Al hacer un inventario de los alcances y usos de la evaluación, comienzo señalando que la evaluación tiene diversos significados para diversas personas, especialmente en un foro internacional con múltiples audiencias. Hace unos veinte años, Glass y Ellet afirmaban que (1980): “La evaluación, más que cualquier otra ciencia, es lo que la gente dice que es, y actualmente la gente dice que la evaluación es muchas cosas distintas”.¹ Aunque perduran acalorados debates sobre la naturaleza y los usos de la evaluación, se discuten — por ejemplo, sus beneficios y usos por grupo objetivo, así como la manera en que debe mejorarse el diseño e implementación de las evaluaciones para garantizar su utilidad — los cambios y el desarrollo en las prácticas de evaluación nos brindan una mayor comprensión sobre lo que sí es la evaluación y lo que no es la evaluación.

Relacionadas, pero no iguales

Primero, una evaluación no es una auditoría. Persiste una confusión general entre estas actividades, en parte porque evaluación es un término común para cualquier actividad de valoración, incluyendo otras formas de investigación aplicada como la investigación-acción o el análisis estratégico, frecuentemente considerados sustitutos. Esta confusión también se explica por el hecho de que algunas oficinas auditoras nacionales han ampliado su campo de acción de la tradicional auditoría financiera del gasto público —con énfasis en la eficiencia y economía— a “sistemas de control administrativo” y de “optimización de los recursos” (VFM, por sus siglas en inglés) que auditan la efectividad en la administración de los recursos. En efecto, la Oficina del Auditor General (OAG) de Canadá nos muestra que un 60% del presupuesto del 2001 es para “auditorías de desempeño” (VFM) que miden la correcta implementación de políticas y programas, con un 33% destinado a “actividades de auditoría de estados financieros” (Sutherland, 2002). Para algunos expertos, este hecho representa un marcado progreso en la regulación de actividades gubernamentales. Esta *explosión de auditoría*, como le llama Michael Power, ha creado un nuevo estilo de rendición de cuentas que se manifiesta en rigurosos y detallados controles con dientes de tiburón. Como recientemente lo describió Onora O’Neil en las BBC Reith Lectures de 2002 (Una Cuestión de Confianza):

“La nueva cultura de rendición de cuentas busca, más que nunca, un perfecto control administrativo de la vida institucional y profesional... La idea original de la auditoría ha sido exportada de su contexto financiero original a los procesos y sistemas no financieros, adoptando la forma de un escrutinio detallado. Los indicadores de desempeño son utilizados para medir, con supuesta precisión, el desempeño adecuado o inadecuado”.

Segundo, las evaluaciones son complementarias pero distintas a los sistemas administrativos con base en resultados que crecientemente se han incorporado a la regulación del sector público (Davies, 1999). La motivación para practicar una administración de desempeño es atribuible en buena medida al crecimiento de la Nueva Administración Pública (NPM, por sus siglas en inglés) durante la década de los años 80, con su propuesta de descentralización de responsabilidades en los programas y servicios, asegurando al mismo tiempo que las instituciones públicas se responsabilicen por los resultados de los programas. La rendición de cuentas con base en resultados requiere que las organizaciones demuestren cómo serán gastados los recursos públicos, articulando a su vez lo que deben alcanzar los programas en términos de resultados, incluyendo indicadores e instrumentos de recolección de datos para medir el cumplimiento de los objetivos y estableciendo estándares de desempeño o puntos de referencia para valorar el progreso de los programas, así como análisis periódicos para la toma de decisiones internas e informes públicos (Horsch, 1996). En resumen, los sistemas de rendición de cuentas con base en resultados responden a tres preguntas: ¿Qué objetivos trata de alcanzar el programa? ¿Qué progresos está teniendo? ¿Se han alcanzado los resultados deseados? Por otro lado, la evaluación de programas responde a preguntas distintas, pero igualmente importantes: ¿por qué está teniendo éxito o está fracasando el programa? ¿Qué resultados inesperados surgieron? ¿Qué cambios podrían ser necesarios para mejorar la efectividad? La administración de desempeño responde entonces al “qué”, mientras la evaluación responde al “cómo” y al “por qué”.

La declaración más reciente del gobierno canadiense (Treasury Board Secretariat, 2001) sobre políticas de evaluación señala que:

- Administrar por resultados es la principal responsabilidad de los administradores públicos... de ellos se espera que definan anticipadamente los resultados deseados, que enfoquen continuamente su atención hacia el cumplimiento de esos objetivos, que midan su desempeño de manera regular y objetiva, y que vayan aprendiendo y adaptándose para mejorar la eficiencia y la efectividad...
- La evaluación puede apoyar a los administradores en sus esfuerzos por rastrear y reportar el desempeño real. También puede ayudar a los tomadores de decisiones a valorar objetivamente los resultados de programas o políticas. Esto diferencia la evaluación de la auditoría interna —es un procedimiento que ofrece garantías en cuanto a la estrategia de administración de riesgos de una agencia o departamento, su marco de control administrativo y la información financiera y no financiera que se utiliza para formular reportes y para la toma de decisiones.

Por consiguiente, se instruye a los departamentos gubernamentales a “incorporar la disciplina de la evaluación en la administración de políticas, programas e iniciativas con el fin de:

- Desarrollar marcos conceptuales para la rendición de cuentas y la administración con base en resultados, a ser aplicados en las políticas, programas e iniciativas nuevas o renovadas;
- Establecer prácticas de seguimiento y de medición de desempeño;
- Evaluar temas relacionados con la implementación y administración de políticas, programas e iniciativas en sus primeras fases; incluyendo aquellas que se dan a través de convenios de asociación (evaluaciones en las fases formativas o intermedias); y
- Evaluar los temas concernientes a relevancia, resultados y costo-efectividad”.

La evaluación de programas, entonces, nos ofrece la información sobre la efectividad requerida en informes tipo PPBS (Sistemas de Planeación, Programación y Presupuestos), con todo y sus controles centrales. Adicionalmente, cuando la evaluación se vincula a la administración con base en resultados (y a su énfasis de permitir que los administradores administren), se consolida un sistema de rendición de cuentas que constituye una poderosa herramienta de administración.

Beneficios y usos de la evaluación

“Ser eficiente era rendir cuentas sobre el gasto; ser efectivo era alcanzar resultados significativos” (Parata, 1998).

Esta cita nos recuerda el divertido episodio de la serie televisiva de la BBC, *Yes Minister*, en la cual el hospital más eficiente del Ministerio no sólo era eficiente sino que también carecía de pacientes. Cuando el Ministro preguntó si eso tenía algún sentido, le respondieron que “los pacientes obstaculizarían sus mediciones de eficiencia de desempeño”. El punto aquí no es desacreditar la eficiencia sino recordar la importancia de los resultados.

En el contexto canadiense, la evaluación se preocupa esencialmente por la efectividad, de lo cual surge una pregunta contrafactual: ¿qué hubiese ocurrido de no existir el programa? Se podrían utilizar métodos de observación o de correlación para demostrar la incidencia de los efectos deseados, y pautas experimentales y cuasi-experimentales para determinar si los efectos observados son atribuibles a la intervención y no a otros factores. Las categorías o temas clave son cuatro: la racionalidad del programa, su continuada relevancia, sus impactos y efectos, y las alternativas.

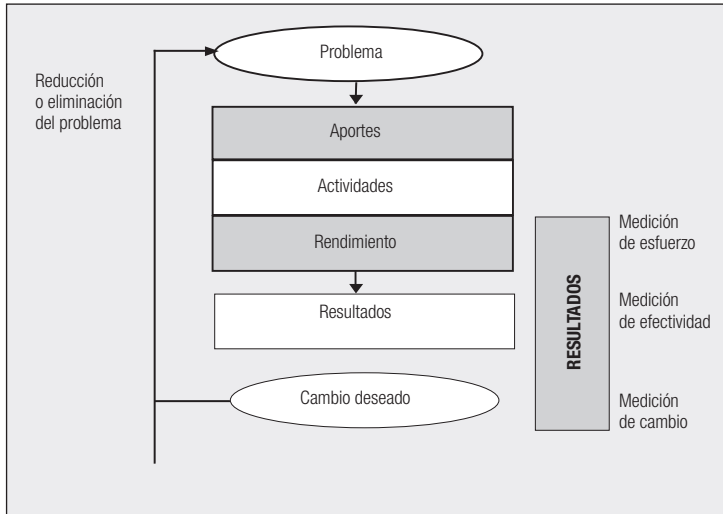
En términos prácticos, una evaluación intenta determinar, como resultado de la implementación de una política, programa o iniciativa: (Fowler, 2002):

- ¿Qué ha cambiado? Identificación de las diferencias.
- ¿Qué ocasionó el cambio? Atribución.

- ¿Es o no es lo que se proponía? Juicio.
- ¿Qué sigue? Aprendizaje y ajustes.

La figura 2.1 presenta un modelo lógico utilizado para mostrar los procesos causales de una intervención y analizarlos como parte de la evaluación.

Figura 2.1 **Modelo lógico típico para evaluar una intervención**



Panorama cambiante

Debemos notar, sin embargo, que los alcances de una evaluación van desde el simple propósito de evaluar la consistencia de una política o programa con los resultados deseados, hasta un enfoque amplio con beneficios múltiples.² Como ejemplo, en 1980 Scriven escribió que: “La evaluación es lo que es. Determinar su mérito o su valor, y para qué se utilizará, es harina de otro costal... Lo malo es malo y lo bueno es bueno y le corresponde al evaluador hacer la distinción”. Weiss contrarresta esta idea afirmando: “El propósito de la evaluación es medir los efectos de un programa con respecto a los objetivos que se propuso cumplir, a fin de contribuir a la siguiente toma de decisiones sobre el programa y a mejorar la programación futura”. Posteriores definiciones reflejan consistentemente un enfoque más amplio, con un énfasis mayor en la variedad de beneficios vinculados a la evaluación. Patton (1997) proporciona, quizás, la definición por excelencia:

“La evaluación de programas es la recolección sistemática de información sobre las actividades, características y resultados de los programas, lo cual permite hacer juicios, mejorar su efectividad o informar para futuras programaciones”.

En otras palabras, las evaluaciones pueden utilizarse para hacer juicios generales, facilitando mejoras y generando conocimientos. En su revisión de la bibliografía pertinente, Chelimsky (1977) concuerda con que casi todos los beneficios especificados son subcategorías de las tres siguientes áreas conceptuales: rendición de cuentas, desarrollo y conocimiento. Vamos a examinar brevemente cada una de ellas.

Rendición de cuentas y toma de decisiones

La evaluación es un apoyo para la rendición de cuentas del desempeño y la ejecución de gastos de un programa. Los sistemas modernos de rendición de cuentas y de evaluación tienen sus raíces en la historia del presupuesto general de gastos del gobierno. En 1971 fue traído a Canadá, desde los Estados Unidos, el Sistema de Planeación, Programación y Presupuesto (SPPP), incluyéndose en el formato de los informes del gobierno federal para los estimados anuales de gastos, si bien el gobierno estadounidense ya había suprimido el SPPP dos años antes. La experiencia estadounidense a partir de 1965 había mostrado que era muy difícil generar información objetiva y que los juicios subjetivos estaban llenando el vacío que se había creado. Ultimadamente, el SPPP centralizaba el poder en el Ejecutivo a expensas del Congreso, por lo que el Congreso prescindió de él.

En Canadá, el legado del SPPP y sus diversas transformaciones cayeron durante el 2002 en el mismo problema intelectual que se había presentado en 1971: ¿Cómo capturar los resultados del programa a modo de satisfacer los requerimientos de la agencia central? ¿Cómo generar enunciados altamente globales sobre los logros del programa que sean verdaderos y verificables? (Sutherland, 2002). La evaluación de programas nace en 1977 como un mandato obligatorio para todos los Ministerios gubernamentales federales, a fin de enfrentar la carga de tener que determinar los impactos de los programas, en ciclos de tres a cinco años. Desde la perspectiva de la rendición de cuentas, la evaluación “se enfrenta con interrogantes de causa-efecto que requieren métodos capaces de relacionar los hallazgos con las intervenciones, de la manera más cercana y concluyente posible” (Chelimsky, 1997).

Jacques Toulemonde (2000, 351) documenta una historia similar en Europa:

“Es sabido que la evaluación nació en los Estados Unidos con el Sistema de Planeación, Programación y Presupuesto (SPPP). El modelo fue importado en los años setenta por la mayoría de los países del norte de Europa en donde se crearon agencias, unidades y comisiones para realizar análisis de políticas. Estas instituciones hacían evaluaciones *ex ante* y *ex post*. Tenían un alcance interministerial y su objetivo era introducir racionalidad científica al proceso presupuestario. Países como el Reino Unido, Holanda y Francia importaron el modelo estadounidense con entusiasmo...”

En la actualidad, todos los países que conocieron este modelo de evaluación lo han abandonado. El cambio más completo y abrupto se dio en Francia, en donde el SPPP fue eliminado a principios de la década de los ochenta. Sin embargo, “el bebé fue expulsado con todo y agua del baño”, a tal grado que la evaluación perdió todo el apoyo gubernamental durante casi diez años.

En otros países la transición fue más progresiva. Por ejemplo, en Holanda el Ministerio de Finanzas realizó un cambio paulatino de la evaluación basada en el SPPP a un sistema más pragmático. En el Reino Unido la Comisión de Análisis de Políticas creada en la década de 1970 se disolvió cuando el gobierno de Thatcher relanzó las prácticas de evaluación con el eslogan de “optimización de los recursos”.

Desde mediados de la década de los ochenta, la difusión de las prácticas de evaluación a nivel europeo se vio fomentada por las redes profesionales de investigación internacional de políticas y ayuda al desarrollo. La OCDE creó grupos de trabajo dedicados a la evaluación, como parte del Comité de Asistencia al Desarrollo (CAD) y las políticas PUMA (*Public Governance and Management*) de administración pública. El Banco Mundial y otras instituciones internacionales como la Organización Internacional del Trabajo también ayudaron a difundir la cultura de la evaluación, ya sea como un requisito para el financiamiento a países en desarrollo o como una herramienta en la administración de proyectos. La Unión Europea, por medio de sus regulaciones a los llamados Fondos Estructurales, jugó un rol más decisivo en el desarrollo de capacidades para la práctica de la evaluación. De acuerdo con las reglas adoptadas en julio de 1993 para el periodo 1995-1999, la evaluación *ex ante* debía ser un prerrequisito para el financiamiento; y el seguimiento y evaluación *ex post*, una obligación de las autoridades regionales y nacionales. La cantidad de evaluaciones en función de estos Fondos Estructurales se quintuplicó en comparación al quinquenio anterior. Además de estos factores externos de fomento, también surgieron fuerzas internas peculiares en ciertos países: presiones por parte del Parlamento en Alemania y en Escandinavia, iniciativas de oficinas nacionales de auditoría en Holanda, Suecia y el Reino Unido, y presiones de los ministerios de finanzas, tal es el caso de Italia (Toulemonde, 2000, 352-353).

Al mismo tiempo, es interesante observar que la demanda de evaluaciones se está desarrollando más rápidamente a nivel europeo y nacional que a nivel regional y local (Leeuw, Toulemonde y Brouwers, 1999):

“Si bien en la actualidad casi todas las políticas europeas están sujetas a evaluaciones periódicas, tan solo un diez por ciento de estas corresponde a políticas regionales que son evaluadas periódicamente en el Reino Unido, Dinamarca, Holanda y Francia”.

Ciertamente, este limitadísimo ejercicio de prácticas de evaluación a nivel subnacional es un desafío para el Programa de Desarrollo Económico y del Empleo a Nivel Local (LEED, por sus siglas en inglés).

Desarrollo

Las evaluaciones también facilitan el desarrollo de políticas y programas, ya que pueden:

- Mejorar el diseño de programas y la articulación clara de objetivos realistas y alcanzables con base en los hallazgos de evaluaciones previas.
- Mejorar la implementación a través de evaluaciones de los procesos del programa.
- Mejorar el costo-efectividad de los programas mediante la investigación de procedimientos alternativos que permitan alcanzar los objetivos al menor costo.
- Apoyar prácticas administrativas eficaces, al validar indicadores y objetivos de desempeño e identificar las prácticas eficaces.
- Desarrollar capacidad analítica y de medición.

Los métodos formativos que comúnmente son utilizados en la evaluación del desarrollo, incluyen: el estudio de casos, la síntesis de investigaciones, las auto-evaluaciones internas, las mediciones de desempeño y el seguimiento. Estas actividades de evaluación de procesos no suelen examinar los resultados hasta que el programa ya está bien establecido.

Conocimientos y destrezas

Como mínimo, la evaluación genera conocimientos sobre las necesidades existentes o potenciales y sobre la programación requerida para satisfacer dichas demandas. Es decir:

- Incrementa los conocimientos sobre las necesidades y problemas.
- Incrementa el conocimiento de las prácticas y programas eficaces.
- Incrementa los conocimientos sobre programación.

La evaluación puede desarrollar capacidades para mejorar el diseño, seguimiento y perfeccionamiento de los programas a través de:

- Aprender a tener un pensamiento más crítico.
- Mejorar las actitudes hacia la evaluación.
- Desarrollar capacidades para utilizar, entender y llevar a cabo evaluaciones.

A diferencia de la evaluación orientada al juicio o al desarrollo, en la que las decisiones suceden a los hallazgos, la evaluación orientada al conocimiento se utiliza para influir en el pensamiento y profundizar la comprensión mediante más conocimientos. En la evaluación orientada al conocimiento, el uso conceptual o iluminador de los hallazgos, más que el instrumental, “puede ser tan específico

como para clarificar el modelo del programa, comprobar la teoría, distinguir entre tipos de intervenciones, resolver cómo medir los resultados, generar lecciones aprendidas o proponer opciones” (Patton, 2001: 332). El personal local del programa adquiere ideas y elementos nuevos mediante estas evaluaciones, lo cual enriquece su comprensión de lo que es el programa y de lo que hace.

Utilidad y humildad

En la actualidad, ya hay una amplia bibliografía sobre los usos y los no-usos de la evaluación. Este crecimiento bibliográfico se dio como respuesta a la crisis de confianza que se generó en los años 70 cuando diversos estudios revelaron que las evaluaciones no eran utilizadas rutinariamente como elemento medular en la toma de decisiones sobre políticas. Subsecuentemente, se ha puesto más atención a los múltiples usos que son parte del diseño y los objetivos de la evaluación. Se postularon las categorías generales de rendición de cuentas, desarrollo y conocimiento, como resultado de investigaciones sobre los “usos alcanzados pero no deseados” en ciertos tipos de evaluación (Chelimsky, 1999). Ciertamente y según mi experiencia, las evaluaciones formativas, prospectivas y acumulativas tienen usos muy diferentes en las distintas etapas del ciclo de desarrollo de políticas. Las evaluaciones prospectivas nos ofrecen un compendio de las evidencias existentes, mediante el meta-análisis o análisis exploratorio de una base de datos sintética. Sus resultados pueden ser relevantes para las etapas iniciales en el análisis de políticas, cuando se identifican problemas y se examinan soluciones alternativas. Las evaluaciones formativas en la implementación de nuevos programas le indican a los tomadores de decisiones si los ingredientes adecuados están efectivamente llegando a su lugar, mientras que las evaluaciones acumulativas miden y explican los resultados del nuevo programa. Lo anterior es una pequeña muestra de los diversos enfoques que pueden darse a una evaluación y demuestra que una evaluación no puede ser de todo para todos.

Sin embargo, hay quienes se quejan del uso diferencial de las evaluaciones, argumentando que se trivializan. Por ello, es importante recordar el propósito central de la evaluación:

“La evaluación consiste en aprender de la experiencia para mejorar el trabajo futuro. Por lo tanto, la evaluación debe tener los mismos objetivos o propósitos que los proyectos, políticas o programas que están siendo evaluados —es decir, *debe crear beneficios sustentables para los grupos objetivo...* Y no podrá hacerlo si no se utiliza. Por lo tanto, todas las evaluaciones deben encauzarse a la utilidad” (Eggers, 1999:93).

“El objetivo de la evaluación no es su utilidad, su objetivo es la mejoría social. El uso, entonces, no es el fin último, sino que es un medio mediante el cual la evaluación consigue la mejora social” (Henry, 2000:3).

Como apunta Chelimsky, en un sentido muy amplio lo anterior expresa efectivamente el fin último de la evaluación. A largo plazo, los beneficios locales de la evaluación contribuirán a una programación social más efectiva, generando ahorros financieros y mejoras en el bienestar. En el corto plazo, sin embargo, hay usos específicos e inmediatos un poco abajo de la cadena.

Además, Eggers reconoce que las evaluaciones “no necesariamente son utilizadas, aún cuando sean utilizables. Una evaluación debe guiar el caballo hacia el agua, aunque no logre hacerlo beber”. Lo anterior admite que, aun con evidencias empíricas convincentes, la evaluación es sólo una consideración entre otras, compitiendo con apreciaciones políticas y de opinión pública. También es cierto que la creciente tendencia a ciclos más cortos en el desarrollo de políticas descarta la posibilidad de llevar a cabo evaluaciones más rigurosas que deben hacerse dentro de los presupuestos y calendarios establecidos, por lo que muchas veces la decisión sobre la política a seguir depende más de evaluaciones impresionistas. A pesar de dichas limitaciones, la evaluación sigue siendo una de las pocas fuentes confiables de evidencia sobre los impactos y la efectividad de las diversas políticas. Para aumentar su utilidad general, la evaluación debe sostenerse como una actividad de importancia a largo plazo y como un surtidor de conocimientos que podría resultar útil en investigaciones específicas.

Más allá de todo lo anterior, existe el problema real de comunicar los resultados de la evaluación con un lenguaje claro y en un contexto apropiado, a fin de que los diseñadores de políticas los tomen en cuenta. Weiss (1998) recomienda que los evaluadores “se esfuercen en la comunicación de sus hallazgos... (pero) sin insinuar expectativas irreales en cuanto a su uso”. Patton (1997), por otro lado, argumenta que las estrategias participativas que involucran a los actores de un programa con el propósito de reclutar sus conocimientos y compromiso, bien pueden incrementar la utilidad de las evaluaciones.

Retos y oportunidades

Soy de la opinión que la evaluación y el aprendizaje sucesivo se producen en un contexto organizado. Para las políticas y los programas de desarrollo económico y del empleo a nivel local, el contexto organizado incluye a los siguientes grupos de potenciales usuarios: patrocinadores (p. ej. gobiernos, organizaciones sin fines de lucro, fundaciones), personal y operadores del programa, usuarios del programa, y entes de la sociedad civil que sean activos en la vida comunitaria y social. Estos grupos son los socios locales que experimentan y proponen innovaciones para la solución de sus problemas económicos y sociales. El reto es crear y aprovechar las oportunidades para desarrollar estrategias de evaluación participativa entre los múltiples actores. La contribución de la evaluación es proporcionar información confiable, precisa y efectiva para la toma

de decisiones. De igual manera, puede ayudar a los socios locales a ponderar y evaluar las evidencias, produciendo un impacto más perdurable que el de un simple hallazgo.

La tarea es considerable si tomamos en cuenta que de acuerdo a las estimaciones, sólo un diez por ciento de las políticas regionales y locales se evalúan en la actualidad. Sin embargo, las evaluaciones de casos a nivel local bien pueden ofrecer un contexto para indicadores más globales. Las evaluaciones síntesis y los meta-análisis que combinan hallazgos de evaluaciones que se basan en la comunidad “para generar patrones y lecciones más sistemáticos constituye cada vez más un modo importante de integrar las evaluaciones descentralizadas al sistema central para ser utilizadas en el diseño de políticas” (Patton, 2000:15). No obstante, si las evaluaciones que se basan en datos a nivel individual se adoptan para comunidades, existe el reto metodológico de convertir las medidas de efectividad individual en hallazgos de impacto para la localidad.

También debemos recordar que las estrategias de evaluación deben adecuarse a las innovaciones en los programas e iniciativas de desarrollo local, en cuyo caso las evaluaciones prospectivas podrían resultar más convenientes que las retrospectivas.

Adicionalmente, las evaluaciones deben reconocer la naturaleza horizontal de las políticas y programas de desarrollo local, la cual comúnmente abarca diversas jurisdicciones. La cohesión social, el capital social y la innovación social son fenómenos difíciles en cuanto a la recolección sistemática de datos y su cuantificación, justificándose por ende un enfoque multidisciplinario.

A pesar de estos desafíos, el potencial de aprendizaje que nos ofrece la disciplina de la evaluación es inmenso; y los conocimientos y visión que se obtienen mediante su uso ciertamente valen la pena.

Notas

1. Citado en Heinecke, Blasi, Milman y Washington, 1999, p. 1.
2. Citado en McGuire (2002).

Referencias bibliográficas

- CANADA Office of the Auditor General of Canada. “Annual Reports of the Auditor General for the Fiscal Years”, 1982-83 (Capítulo 3), 1992-93 (Capítulos 8, 9, 10), 1995-96 (Capítulo 3), 2002-02 (Capítulo 6). Disponible en www.oag-bvg.gc.ca
- CANADA Treasury Board. Evaluation Policy, Ottawa, febrero 2001. Disponible en www.tbs-sct.gc.ca/pubs_pol/dcgpubs/tbm_161/ep-pe_e.html

- CHELIMSKY, Eleanor, "Thoughts for a New Evaluation Society", *Evaluation*, vol. 3 (1): 97-118.
- DAVIES, Ian C., "Evaluation and Performance Management in Government", *Evaluation*, vol. 5 (2): 150-159.
- EGGERS, Helmut W. y Eleanor Chelimsky, "Purposes and Use: What Can We Expect?" *Evaluation*, vol. 5 (1): 92-96.
- FOWLER, Alan, "Measuring Non-tangible Outcomes: The Challenge for Impact Assessment", Australia, abril 2002, disponible en www.mande.co.uk/docs/fowler%20presentation.pdf.
- HEINECKE, W. F., Laura Blasi, Natalie Milman y Lisa Washington, "New Directions in the Evaluation of the Effectiveness of Educational Technology", The Secretary's Conference on Educational Technology, 1999. Disponible en www.ed.gov/technology/techconf/1999/whitepapers/paper8.html
- HENRY, Gary T., "Using Evaluation Findings for Policy: A Realist Perspective", artículo presentado en la Cuarta Conferencia de la Sociedad Europea de Evaluación (ESS), Lausana, 12-14, octubre 2000.
- HORSCH, Karen, "Results-based Accountability Systems: Opportunities and Challenges", Harvard Family Research Project, *The Evaluation Exchange*, vol. II, núm. 1, invierno 1996. Disponible en <http://gseweb.harvard.edu/~hfrp/eval/issue3/horsch.html>
- KVTASTEIN, Olav A., "Extracting Lessons from Central Debates in order to Improve Policy Implication of Evaluations", artículo presentado en la Cuarta Conferencia de la Sociedad Europea de Evaluación, Lausana, 12-14 octubre 2000.
- MACKAY, Keith (ed.), *Public Sector Performance — The Critical Role of Evaluation*. Banco Mundial, Washington, D. C., 1998.
- MCGUIRE, Martha, Literature Review on the Benefits, Outputs, Processes and Knowledge Elements of Evaluation. *Canadian Evaluation Society Project in Support of Advocacy and Professional Development*, Ottawa, octubre 2002.
- O'NEILL, Onora, BBC Reith Lectures 2002 — A Question of Trust. Disponible en www.bbc.co.uk/radio4/reith2002
- OCDE, "Evaluation Feedback for Effective Learning and Accountability", Development Assistance Committee (DAC), Working Party on Aid Evaluation, París, 2001.
- , *Improving Evaluation Practices: Best Practice Guidelines for Evaluation and Background Paper*, PUMA, París, 1999.
- PARATA, Hekia, "The Outputs/Outcomes Nexus: the relative responsibilities between Ministers and the Public Service", en *Lifting the Game from Outputs to Outcomes: Proceedings, Public Service Senior Management Conference*, Comisión de servicios estatales, Wellington, enero, 1998.
- PATTON, Michael Q., *Utilization-focused Evaluation: The New Century Text*, Sage, California: Thousand Oaks, 1997.
- , "A Vision of Evaluation that Strengthens Democracy", discurso presentado en la Cuarta Conferencia de la Sociedad Europea de Evaluación, Lausana, 2000.
- , "Evaluation, Knowledge Management, Best Practices, and High Quality Lessons Learned", *American Journal of Evaluation*, 2001, vol. 3 (1): 329-336.

- SUTHERLAND, S.L., *The Office of the Auditor General of Canada: Government in Exile?* Documento de Trabajo 31, School of Policy Studies, Queen's University, Kingston, Ontario, septiembre 2002.
- TOULEMONDE, Jacques, "Evaluation Culture(s) in Europe: Differences and Convergence between National Practices", *Quarterly Journal of Economic Research*, 3/2000, S.350-357. Disponible en www.diw.de/english/publikationen/vierteljahrshefte/jahrgang00/content_3.html
- UNITED STATES, *Case Study Evaluations*. General Accounting Office, Washington, D. C. noviembre, 1990.
- WEISS, Carol H., "Have We Learned Anything New About the Use of Evaluation?" *American Journal of Evaluation*, invierno, 1998, vol. 19 (1).

Capítulo 3

Evaluación: evidencia para la política pública

Por
Robert Walker,
Profesor de Política Social,
Universidad de Nottingham, Reino Unido

Esta contribución provee una perspectiva general de los desafíos políticos, institucionales y metodológicos que enfrenta la evaluación de las políticas públicas con la visión de estimular una respuesta constructiva y de colaboración mutua. Comienza especificando las diferentes interrogantes que surgen en la evaluación de las políticas y las relaciona con las técnicas que las responden. La segunda sección se enfoca en las amenazas a una evaluación efectiva de las políticas que surgen del proceso, naturaleza y mercado de la evaluación de políticas. La última sección versa sobre los retos de construir y conformar una cultura de evaluación vital y sostenible. Se utilizan ampliamente los estudios de caso, tanto de evaluaciones de políticas específicas, como de ejemplos del Reino Unido donde, desde la elección del nuevo gobierno laborista en 1997, se ha buscado introducir una cultura de evaluación de políticas en el centro del diseño de las políticas del gobierno central.

“¿Funcionará esta política?” “¿Qué clase de políticas han funcionado en el pasado?” Si preguntas como éstas pudieran responderse satisfactoriamente, se reduciría la mayor parte del elemento de riesgo en las políticas y la formulación de políticas se convertiría en una ciencia.

Sin embargo, la realidad se sitúa muy lejos de tal aspiración. No sólo esta clase de preguntas se formulan raramente en las ciencias sociales, sino que los científicos sociales son incapaces de responderlas con un cierto grado de certeza. Con gran esfuerzo, se logran obtener respuestas sólidas a preguntas muy específicas que limitan la posibilidad de generalizar. En tanto que, las grandes cuestiones estratégicas generan respuestas imprecisas matizadas por calificativos. En consecuencia, la evidencia en la evaluación está destinada a ser sólo una de las tantas fuentes de información utilizadas por los diseñadores de políticas y ésto continuará siendo una empresa riesgosa. No obstante, la evaluación hace a las políticas menos aburridas y podría ser, de hecho, bueno para la salud de la democracia.

Sin embargo, reducir el riesgo de introducir políticas deficientes o rechazar políticas apropiadas, incluso por un breve margen, podría prevenir que millones de dólares, euros o libras fueran desperdiciados, permitiendo que estos recursos se sumen al bienestar de la humanidad. Por lo tanto, corresponde a la comunidad encargada de las políticas buscar mejorar la calidad, cantidad y uso de la evidencia

en la evaluación. Esto necesitará una estrecha alianza entre aquéllos que buscan la evidencia en la evaluación —quienes necesitan pedir la información con suficiente anticipación, ayudar a financiar su obtención y estar preparados para actuar, tanto frente a las conclusiones favorables, como a las que no lo son— y quienes la producen —que deben comprometerse con las prioridades de los diseñadores de política, mientras que protegen y desarrollan estándares para la investigación científica—. Ésto requiere de instituciones que promuevan, apadrinen y faciliten esta alianza, fomentando el diálogo productivo y asegurando que las diferentes recompensas obtenidas del trabajo realizado sean compartidas adecuadamente.

Esta perspectiva general de los temas comienza especificando diferentes preguntas hechas en la evaluación de políticas y las relaciona con las técnicas para responderlas. La segunda sección se enfoca en las amenazas o una evaluación eficiente de las políticas que se origina en el proceso, naturaleza y mercado de la evaluación de políticas. La sección final trata acerca de los desafíos de construir y conformar una cultura sostenible y vital de evaluación. Se utilizan ampliamente los estudios de caso, tanto de evaluación de políticas específicas como de ejemplos del Reino Unido donde, desde la elección del nuevo gobierno laborista en 1997, se ha buscado introducir una cultura de evaluación de políticas en el centro del diseño de las políticas del gobierno central.

Una pregunta sobre evaluación

Hay distintos tipos de evaluación de políticas pero una forma simple de categorizarlos es en términos de la formulación pregunta y cuándo se formula (Tabla 3.1). Existen dos preguntas básicas de evaluación, una descriptiva: “¿funciona la política?”, y la otra, analítica: “¿por qué?”. Sin embargo, para evitar las implicaciones teleológicas de la pregunta del “¿por qué?”, es preferible reformular la segunda pregunta a el “¿cómo?”: “cómo funciona y cómo no funciona la política.” Las evaluaciones relacionadas con la primera pregunta reciben el término de evaluaciones “acumulativas”, “de programas” o “de impacto” y son típicamente cuantitativas (Orr, 1998; Greenberg y Schroder, 1997; Worthen *et al.*, 1996). Las evaluaciones que se enfocan en la segunda pregunta son denominadas frecuentemente “formativas” y son a menudo cualitativas (Patton, 2002; Pawson y Tilley, 1997; Yanow, 1999). Sin embargo, se ha utilizado una sobreabundancia de términos diferentes para describir a la evaluación formativa, que refleja sutil y algunas veces radicalmente, diferentes posiciones ontológicas (véase adelante).

De manera creciente, las evaluaciones de políticas públicas combinan elementos, tanto de la evaluación acumulativa como de la evaluación formativa (Gibson y Duncan, 2002). El hecho de que éste sea comparativamente un desarrollo reciente, en especial en los Estados Unidos, es de alguna forma un tanto

Tabla 3.1 Preguntas de evaluación

Perspectiva temporal	Pregunta de evaluación	Método de evaluación característico	Contraparte formativa de la pregunta de evaluación	Entornos característicos de evaluación formativa	Ejemplos
Pasado de largo plazo	¿Qué funcionó?	Meta-análisis revisión sistemática	¿Cómo funcionó?	Revisión sistemática	Ashworth <i>et al.</i> (2002) (Caso de estudio 4)
Pasado	¿Funcionó la política?	Evaluación retrospectiva	¿Cómo funcionó o no?	Entrevistas retrospectivas Juicios participativos (estudios cognoscitivos) estudios de caso retrospectivos	Huby y Dix (1992) (Caso de Estudio 3)
Presente	¿Está funcionando la política?	Monitoreo • Series de tiempo interrumpidas • Experimentos naturales	¿Cómo está funcionando o no funcionando?	Estudios de Proceso Implementación de la evaluación Etnografía	
Presente a futuro	¿Existe algún problema?	Investigación básica análisis de políticas	¿Cuál es el problema?	Investigación Básica Reconocimiento rápido	
Futuro cercano	¿Podemos hacer que esta política funcione?	Prototipos micro-simulación	¿Cómo podemos hacer que esta política funcione?	Teoría del cambio Investigación participativa Investigación de acción	Loumidis <i>et al.</i> (2001) (Estudio de caso 2)
Futuro	¿Funcionará esta política?	Evaluación del programa (evaluación de impacto o acumulativa) • Asignación aleatoria • Diseños de emparejamiento • Diseños de cohorte • Controles estadísticos	¿Cómo funcionará / o no funcionará?	Teoría del cambio Laboratorio de evaluación	Michalopoulos <i>et al.</i> (2002) (Estudio de caso 1) Hills <i>et al.</i> (2001) (Estudio de caso 8)
Futuro de largo plazo	¿Qué política funcionaría?	Evaluación prospectiva • Micro-simulación • Experimentación de laboratorio • Juegos	¿Cómo funcionaría?	Laboratorio de evaluación Consulta Delphi Juegos	Brewer <i>et al.</i> 2001 (Estudio de caso 5) Voyer <i>et al.</i> , 2002 (Estudio de caso 6) Walter <i>et al.</i> (1987) (Estudio de caso 7)

Fuente: OCDE.

desconcertante, ya que uno hubiera pensado cómo natural el querer saber por qué una política funcionó o no. Quizás los evaluadores pensaron que era suficiente saber que una política funcionaba, porque de esta manera podía continuar o implementarse en cualquier otra parte. (El término “proyecto demostrativo” usualmente utilizado en los Estados Unidos para describir evaluaciones de programas, sugiere que los formuladores de política tenían tan gran fe en el paquete de políticas evaluado, que la probabilidad de fracaso y la necesidad de analizar el porqué era rara vez investigado.) Una explicación alternativa señala que las metodologías diseñadas para responder a las dos preguntas fueron desarrolladas en diferentes ámbitos de las ciencias sociales, y que se unieron tardíamente por necesidad o en reconocimiento al valor de lo interdisciplinario en la investigación aplicada de las políticas (Ritchie y Lewis, 2003).

La otra dimensión para clasificar los estudios de evaluación se relaciona con el cuándo se formularon las dos preguntas de evaluación en el ciclo de aplicación de la política. El tiempo verbal utilizado en las preguntas de evaluación (presente, pasado o futuro) indicará, en términos generales, si la evaluación se realiza en forma simultánea, en retrospectiva o prospectivamente.

Sin embargo, antes de discutir las implicaciones de estas distinciones, existe una pregunta previa: “¿Cuál es el problema u oportunidad que requiere una respuesta de política institucional?”

Los temas aparecen en la agenda de políticas por una gama de razones. Éstas incluyen: la ocurrencia de una crisis; como resultado de un cambio temporal en lo económico, lo político o lo social; como producto de las actividades exitosas de individuos, grupos de interés o directores motivados; y por errores y equivocaciones causados por los políticos o administradores (Hall et al., 1975; Guess y Farnham, 2000). Cuando surge un tema, que podría justificar una respuesta de política pública, el consejo que se ofrece en la mayoría de los manuales de política es que debería llevarse a cabo una investigación básica para delinear la naturaleza del problema e identificar, cuando sea posible, caminos de causalidad que podrían indicar puntos dónde intervenir a través de las políticas (Cm., 1999). Sin embargo, los manuales son frecuentemente ignorados y esta etapa, en el proceso de formulación de las políticas, es con frecuencia omitida o realizada superficialmente. Cuando se piensa que una cuestión es importante y urgente, las políticas a menudo son ideadas e implementadas antes que el tema sea entendido a cabalidad. Aun cuando éste no sea el caso, la investigación realizada normalmente se interpretará como investigación aplicada o análisis de política en vez de evaluación. Sin embargo, en tanto que muchas políticas son iatrogénicas, una respuesta al fracaso de una política preexistente, está en la investigación preliminar sobre sus alcances: “Si la política actual no funcionó, ¿cuáles fueron las razones?” Además, si la evaluación alguna vez se convierte en un elemento central en el proceso de las políticas, la evidencia de

la evaluación será, en cambio, fundamental para cualquier revisión prospectiva de la política.

Evaluación de programas

Es apropiado comenzar la discusión de los diferentes tipos de evaluación con el modelo en el cual una política es evaluada antes de su implementación total, dado que esto se presenta algunas veces como el tipo ideal de estrategia de evaluación (Orr, 1998). En este caso, la pregunta de evaluación es correctamente expresada en futuro simple: “¿Funcionará esta política?”, aunque el tiempo presente: (“¿Funciona esta política?”) se utiliza algunas veces, lo cual tiene el efecto injustificado de cambiar una pregunta específica por otra general que pareciera no estar delimitada ni en el tiempo ni en el espacio.

Abordar la pregunta “¿Funcionará?”, usualmente implica llevar a cabo una evaluación del programa o un experimento de la política (Greenberg y Schroder, 1997). Conceptualmente ésta es la forma más directa de evaluación. Algunas personas son sujetas a la intervención de la política mientras que otras no lo son y lo que interesa es observar el resultado en ambos grupos. Cualquier diferencia en resultados que se dé entre ambos se interpreta como una medida del impacto de la política, asumiendo que todos los demás factores permanecen constantes. La evaluación de programas se inició en los Estados Unidos con los celebrados experimentos del impuesto negativo al ingreso de principios de los años 70 y se han mantenido como el pilar de la industria de la evaluación de políticas en ese país.

Aun así, existen cuestiones complejas sobre la definición e implementación. Primero, los objetivos de la política necesitan definirse y priorizarse con precisión. (Esto puede ser en sí mismo un incentivo para el mejoramiento en la creación de políticas, que comúnmente se basan en aspiraciones políticas en lugar de objetivos.) Priorizar los objetivos es necesario, ya que los diseños de evaluación rara vez pueden medir el desempeño de múltiples objetivos de la política con igual precisión, y por lo tanto, usualmente se diseñan con respecto del objetivo más importante.

En segundo lugar, debe existir un acuerdo sobre el grado de cambio que constituiría un éxito. Por ejemplo, una aspiración de reducir la tasa de pobreza tiene que ser acompañada por una declaración que indique el número de puntos porcentuales en los que se reducirá la pobreza. Esto es necesario ya que la muestra utilizada en la evaluación debe ser lo suficientemente grande para determinar con adecuada precisión si la política ha sido o no exitosa.

Tercero, algún modelo o teoría del cambio debería llevarlo a uno a suponer que la política que está siendo implementada produciría el cambio anticipado. Tal modelo permitiría diseñar indicadores correctos de resultados. Además,

mientras mejor especificado esté el modelo, mayor será el número de resultados intermedios que pueden incorporarse a la evaluación a fin de valorar el modelo de cambio y generar indicadores de diagnóstico cuando los resultados no concuerden con las expectativas.

Finalmente, es necesario definir la situación contrafactual: qué se habría producido si la política a ser evaluada no se hubiera implementado. La situación contrafactual proporciona una base contra la cual se puede evaluar el desempeño de la política. Por lo general, es inadecuado comparar simplemente el patrón de resultados antes y después que la política se introduce, ya que otras características del contexto de la política pueden cambiar e influenciar la efectividad de la política. Por ejemplo, en una economía próspera es más probable que se reduzca la pobreza aun en ausencia de políticas anti-pobreza; en tales circunstancias, “otras cosas no permanecerían constantes” como lo requiere el modelo de evaluación. La función de la situación contrafactual es la de eliminar los efectos de esos otros cambios para aislar el impacto específico de la política.

El método que separa el impacto de una política con un sesgo mínimo y la precisión máxima supone asignar aleatoriamente a miembros del grupo objetivo de la política a uno de dos subgrupos: el llamado “grupo de acción” que tiene acceso a la nueva política y el “grupo de control” que no lo tiene. Cualquier diferencia observada en los resultados de estos dos grupos puede atribuirse con certeza a los efectos de la política, dado que todos los demás cambios, debido al proceso de aleatoriedad, influirán al azar en ambos grupos.

A pesar de que a menudo es considerada como la regla de oro en la evaluación, la aleatoriedad tiene limitaciones (Bottomley y Walter, 1996; Stafford et al., 2002). Una es la dificultad para asegurar un consenso político. Los políticos usualmente objetan la asignación aleatoria por dos razones interrelacionadas. La primera es que la experimentación niega a ciertas personas el acceso a lo que puede ser considerado como un bien evidente. En realidad, por supuesto, la razón por la que se evalúa una política es la incertidumbre sobre sus beneficios, o por lo menos, el costo-efectividad de la misma. La segunda razón es que el supuesto bien es asignado aleatoriamente, en lugar de asignarlo atendiendo a la necesidad o a la probabilidad de lo que la gente obtendrá de él. Dado que podría ser imposible afirmar por adelantado quién se beneficiará más de la intervención de la política, podría decirse que la asignación aleatoria es un método tan bueno como cualquier otro al asignar recursos escasos. No obstante, se ha probado que estas preocupaciones son los mayores obstáculos para el uso de la asignación aleatoria en Gran Bretaña.

Existen además limitaciones técnicas en cuanto a la asignación aleatoria. La más importante surge cuando se pretende que la política a ser evaluada, afecte tanto al sistema en su conjunto, como a los individuos que lo conforman, o cuando surgen consecuencias no previstas de la política (Bottomley y Walter, 1996).

Considérese, por ejemplo, una política que proporciona a los beneficiarios de la seguridad social un vale que permite a sus empleadores reducir los costos de su contratación. En un experimento que involucró el uso de la asignación aleatoria, sólo una proporción de los trabajadores que cuenta con seguridad social recibirá el vale y por lo tanto gozará de una ventaja competitiva frente a otros que buscan empleo, ventaja que desaparecería si a todos los trabajadores que buscan empleo se les diera el vale. Este efecto, el así llamado sesgo de hacer cola, consiste en exagerar la aparente efectividad de una iniciativa política. Sin embargo, el sistema amplía las consecuencias, por ejemplo, una reducción en la tasa de salarios es más probable que sea subestimada, en tanto, no todos los beneficiarios de la seguridad social reciben un vale: el efecto del equilibrio parcial.

Desafortunadamente, no existe una manera práctica para determinar la escala del sesgo de hacer cola o del efecto del equilibrio parcial (Burtless y Orr, 1986). Los diseños cuasi-experimentales, en los cuales una política puede ser implementada completamente en una jurisdicción y no en otra, pueden permitir la estimación de los efectos de la ampliación en el sistema, mientras que también evitan la necesidad de permitir el acceso a políticas sobre una base aleatoria. Sin embargo, debido a que dos áreas nunca son idénticas, el control de los factores exógenos a la política en sí misma es mucho más débil.

Otras dificultades asociadas con la asignación aleatoria, incluyen el asegurar que el personal encargado de la implementación de la política efectivamente conceda acceso a los recursos al usuario de manera aleatoria y que impida que los miembros del grupo de control obtengan clandestinamente los mismos servicios que el grupo de acción.

Estudio de Caso 1

Evaluación de Programas: El Proyecto de Autosuficiencia - Michalopoulos et al., 2002

Hace una década, el Departamento Canadiense de Empleo e Inmigración decidió investigar el efecto que la estrategia “hacer que el trabajo pague” tendría en la habilidad de los beneficiarios de largo plazo de la seguridad social para hacer la transición a empleos de tiempo completo. Por consiguiente, el Departamento encargó un proyecto de demostración de diez años de duración como parte de una iniciativa de política especialmente diseñada basada en un generoso suplemento temporal al ingreso, el Proyecto de Autosuficiencia (PA). El programa incluía a 9000 familias de padres solos en dos provincias: New Brunswick y British Columbia.

Para medir el impacto del PA, se realizó un experimento social basado en la asignación aleatoria. Se tomó una muestra de los beneficiarios de la seguridad social que tenían más de un año de serlo. Posteriormente, se asignó aleatoriamente la mitad de la muestra al programa y se les ofreció el suplemento

Estudio de Caso 1 (cont.)

PA, mientras que el resto constituyó el grupo de control. Los miembros del grupo del programa podían recibir el suplemento por un máximo de tres años.

Aproximadamente 6500 beneficiarios seleccionados fueron visitados en su hogar para participar de la encuesta base de 30 minutos. Se les dijo a los beneficiarios que habían sido seleccionados para unirse al estudio e invitados a firmar un formulario de consentimiento, después de haber escuchado acerca del programa y del principio de la asignación aleatoria. El índice de respuesta para la encuesta base fue de cerca del 90%. Los encuestados fueron entrevistados 36 y 54 meses después de la asignación aleatoria, y se utilizaron los archivos administrativos de distintos departamentos gubernamentales para monitorear su progreso. La probabilidad de ser asignado al grupo del programa era del 50:50 en las dos provincias, exceptuando un periodo de doce meses en New Brunswick donde se asignó en proporciones iguales a los beneficiarios en tres grupos: un grupo al programa, otro al grupo de control y un tercero al grupo plus PA, en el que se les ofreció orientación y asistencia adicional en la búsqueda de trabajo. Se incorporaron tres estudios adicionales en el diseño: el estudio de quienes solicitaron el Proyecto de Autosuficiencia, el estudio sobre el Proyecto de Autosuficiencia Plus (PA Plus) y el estudio "Cliff". El primero adoptó un diseño experimental que implicaría asignar aleatoriamente a cerca de 3000 nuevos solicitantes de los beneficios de la seguridad social al grupo del programa, permitiéndoles recibir el suplemento doce meses después de la solicitud o de pertenecer al grupo de control. El estudio PA Plus analizó la experiencia de los individuos a quienes se ofreció el PA Plus, lo que permitió la comparación, tanto con el grupo de control como con el grupo que recibió el PA básico, una antes de vencer el periodo de elegibilidad. El estudio Cliff examinó las consecuencias del retiro del suplemento después de haberlo recibido por tres años. Usando historiales administrativos, se siguió la trayectoria de 378 personas identificadas en la encuesta de seguimiento realizada a los 54 meses, al acercarse a la finalización del derecho al suplemento. Una submuestra de 52 participantes de este grupo fueron reclutados para participar en un estudio cualitativo, que constaba de un grupo focal inicial, seguido por tres entrevistas telefónicas; una antes de vencer el periodo de elegibilidad, una a los cuatro meses posteriores a la fecha de vencimiento y otra, ocho meses después.

La evaluación demostró que el PA aumentó el empleo, las ganancias y el ingreso y redujo la utilización de la asistencia social y la pobreza: los miembros del grupo del programa recibieron en promedio \$ 6300 más en ingreso total, incluyendo los pagos de asistencia social durante el periodo de seguimiento de 54 meses. Al combinar el suplemento con los servicios (PA Plus) se ayudó a las personas a encontrar empleos más estables que sus contrapartes en el grupo de control. Los beneficios sociales del PA superaron el costo para el gobierno.

Cuando la aplicación de lo contrafactual es imposible, el escenario contrario debe definirse de otras maneras. Numerosos diseños creativos se han desarrollado (Orr, 1998). Estos incluyen: el diseño de grupos de emparejamiento, donde miembros del grupo de control son escogidos para que sean lo más semejante posible a los miembros del grupo de acción (Loumidis et al., 2001; Brice, 1996), diseños de cohorte en los que se busca aprovechar situaciones en las cuales generaciones sucesivas han seguido la misma trayectoria (Smith et al., 2001), diseños de series de tiempo interrumpidas que pueden ser utilizadas cuando existen series estadísticas previas a la implementación de la política y varias formas de experimentación natural (Curington, 1994). Además, se han diseñado estrategias analíticas, siendo tal vez la más prometedora es el emparejamiento de resultados, que intenta definir una situación contrafactual a partir de la variación en un diseño no aleatorio. Sin embargo, Cullen y Hills (1996, p. 14) argumentan que los diseños cuasi-experimentales en conjunto con el modelado estadístico son “la única solución práctica al sueño irrealizable de una aleatoriedad completa”.

Prototipos

El prototipo de política se orienta a una cuestión muy diferente sobre la evaluación de programas. Cuando se ha tomado en principio la decisión de implementar totalmente una política, la pregunta es: “¿cómo podemos hacer que esta política funcione?” La tarea en el prototipo es, por lo tanto, afinar el contenido de la política para lograr el mejor efecto y determinar la manera óptima de implementación. Estos objetivos ponen un menor énfasis en la medición de los resultados y uno mayor en comprender el proceso de implementación, obteniendo como resultado que la metodología es tanto ecléctica como variada e incluye análisis de trabajo-tarea del tipo emprendido por los investigadores operativos y a gran escala, y a las evaluaciones multimetodológicas con componentes tanto cuasi-experimentales como acumulativos y formativos. La confianza en los datos administrativos es con frecuencia fuerte. La mayoría de las grandes evaluaciones, las así llamadas “piloto”, encargadas por el gobierno laborista británico desde 1997 son llamados con mayor precisión prototipos en vez de evaluaciones de programas.

Estudio de Caso 2

Prototipo: Nuevo pacto para personas con discapacidad- Loumidis et al., 2001

El programa piloto de consejeros personales del Nuevo Pacto para Personas con Discapacidad es bastante típico del diseño de las evaluaciones de política encargadas por el gobierno laborista británico después de 1997. El objetivo era determinar si los receptores de los Beneficios por Incapacidad y de algunos otros beneficios a los discapacitados responderían a una invitación para una entrevista enfocada al trabajo con un asesor personal y si con la asistencia de asesores personales, más personas discapacitadas podían asegurar y retener un empleo remunerado.

Se establecieron dos de seis grupos piloto en un rango de diferentes mercados laborales, el primer grupo piloto fue dirigido por el Servicio de Empleo y el segundo por asociaciones de diversas organizaciones públicas, privadas y voluntarias. El primer conjunto de grupos piloto se inició antes de encargar la evaluación. El diseño no permitió una asignación aleatoria, debido a que el Servicio de Asesores Personales estaba disponible para todas las personas con discapacidad, que eran elegibles en las áreas piloto. Por lo tanto, la invitación a la oferta sugirió el establecimiento de doce áreas de comparación. De hecho, el Departamento también tenía aspiraciones de generar estadísticas de base contra las cuales validar el impacto de cualquier implementación a nivel nacional. Por lo que encargó un diseño que incluía, además de las entrevistas a los solicitantes del programa del Nuevo Pacto, entrevistas a personas con discapacidad seleccionadas de una muestra a nivel nacional seleccionadas fuera de las áreas piloto, pero que habían sido estratificadas de acuerdo a los mismos criterios de selección utilizados en las áreas piloto. La intención era que esta encuesta nacional fuera utilizada para establecer la situación contrafactual frente a la cual se evaluaría la efectividad del programa de Asesores Personales. Un programa completo del proceso de evaluación acompañó el análisis de impacto.

Un aspecto crítico en el diseño de todas las evaluaciones de política es el tamaño anticipado de cualquier efecto. Cuándo el tamaño de las muestras grandes puede ser comparativamente pequeño. Sin embargo, cuando el efecto real resulta ser más pequeño que el esperado, la fortaleza y la precisión de un diseño debe de ser comprobado rigurosamente. En el caso del programa piloto del Nuevo Pacto para Personas con Discapacidad, las restricciones de recursos limitaron los tamaños de las muestras que podían conseguirse (aproximadamente 3000 en total), mientras que la amplitud del esquema demostró ser más pequeño que lo esperado.

Aunque, los estudios piloto para el Nuevo Pacto para Personas con Discapacidad se encargaron explícitamente de proporcionar información a las decisiones sobre las posibilidades de una subsecuente implementación a nivel nacional, la intención fue siempre que tales decisiones se tomaran a mitad del periodo piloto de dos años y antes que los resultados del análisis de impacto estuvieran disponibles. En tales circunstancias, el avance de la política no se benefició enteramente de la evaluación sofisticada o, al menos, no en el corto plazo.

El diseño de prototipos está comúnmente modelado por la rápida velocidad con la cual son implementados, por las expectativas de ofrecer información y por su cercanía al proceso de desarrollo de la política. Los reportes provisionales y periódicos son normales y comúnmente están vinculados a una etapa de inicio de la política en la cual los resultados de la evaluación son utilizados para modificar la implementación en una manera semejante que recuerda a la investigación de acción. Efectivamente, en el Reino Unido la implementación total de la política comúnmente precede a la disponibilidad de los resultados de la evaluación del prototipo (Walter, 2001).

En Gran Bretaña, algunos prototipos se iniciaron con una limitada prescripción del contenido de política. Las agencias ejecutoras fueron las encargadas de desarrollarlo dentro de una estructura documental y los informes descriptivos de las políticas desarrolladas reportados a los diseñadores de política con o sin comentarios sustentados en la investigación (Walter, 2000a).

Seguimiento y evaluación retrospectiva

Cambiar la pregunta de evaluación del futuro al presente, “¿Está funcionando la política?” o al pasado, “¿Funcionó la política?”, significa que definir una situación contrafactual es enteramente imposible por la simple razón de que la política está ya disponible para todos como resultado de su total implementación.¹ En estos escenarios, los evaluadores recurren al seguimiento y a la evaluación retrospectiva.

El seguimiento y la evaluación retrospectiva continúan siendo enfoques populares a pesar de sus obvias limitaciones. Esto se debe parcialmente a que estas evaluaciones forman parte del proceso normal de auditoría de la política, que busca establecer quién ha recibido el servicio y a qué costo, a menudo tomando como referencia a la información administrativa. La evaluación retrospectiva puede ser desencadenada por la sospecha, iniciada a menudo por el seguimiento, de que la política no está funcionando bien. Estos tipos de evaluación no requieren del mismo nivel de compromiso institucional hacia la evidencia basada en la formulación de la política como en el caso de la evaluación de programas. No se localizan, por ejemplo, en la ruta crítica de la idea de la política que en la introducción de la política se demande que los diseñadores de política refrenen su entusiasmo en la implementación para esperar el resultado de una evaluación prolongada. El seguimiento y la evaluación retrospectiva son, por lo general, más baratas que la evaluación de programas, pero tienden a responder a diferentes preguntas en diversas maneras.

La ausencia de una caracterización de una situación contrafactual estable comúnmente significa que se da mayor énfasis a los insumos de recursos y a su conversión hacia la provisión de servicios (eficiencia administrativa), que a la contribución en la entrega de servicios para el cumplimiento de los objetivos de

la política (eficacia). Éste es el territorio de los investigadores operativos, estadistas oficiales y sociólogos con un interés en las instituciones y en la implementación de las políticas, más que de los economistas quienes, especialmente en los Estados Unidos de Norteamérica, han liderado el desarrollo en la evaluación de programas. Frecuentemente, los resultados del seguimiento llegan a ser de dominio público como un compendio de estadísticas discretas (Cm., 2002), más que como evaluaciones globales de políticas específicas, que son, sobre todo, el dominio de las evaluaciones retrospectivas.

Las metodologías empleadas en las evaluaciones retrospectivas tienden a ser eclécticas y bajo circunstancias adversas pueden estimular la creación de diseños creativos. Los enfoques plurales frecuentemente utilizados, en los cuales las experiencias y las opiniones son los actores clave en la implementación de la política, son confrontados y triangulados en la búsqueda de un juicio general sobre la efectividad de la política. Las encuestas basadas en entrevistas personales pueden ser utilizadas para solicitar a los receptores de la política sus puntos de vista, mientras que las entrevistas cualitativas realizadas con los administradores y otros grupos de interés y las técnicas de observación se utilizan en el punto de entrega de los servicios. Estos informes pueden proveer evidencia irrefutable sobre la eficiencia o ineficiencia en la implementación y proporcionar una base firme para las reformas.

Un enfoque particular se da a menudo en los objetivos, ya que un caso a *prima facie* puede usualmente ser construido, porque si una política no está alcanzando a su población objetivo es improbable que sea particularmente eficiente. Dos aspectos son importantes: primero, la proporción de beneficiarios de la política que reciben los servicios innecesariamente debido a un diseño deficiente de la política, la mala administración o fraude, y segundo, la proporción de la población elegible que actualmente recibe el servicio y el uso que le da (Knapp, 1984; Walter, 2004). El primer aspecto puede caracterizarse al recabar estadísticas administrativas, especialmente si los juicios acerca de quien recibe el servicio “innecesariamente” son emitidos en relación con las especificaciones del programa, más que en relación con los resultados de la política (lo que requeriría de la definición de una situación contrafactual). Las especificaciones acerca del uso a menudo plantean dificultades mayores ya que quienes son elegibles pero no lo solicitan son usualmente invisibles a la administración y difíciles para hacer el seguimiento empírico.

Cuando las evaluaciones retrospectivas han buscado evaluar el impacto de la política, usualmente utilizan uno o más de tres enfoques: análisis de tendencias, cuasi-experimentación y reportajes. El primer enfoque implica la búsqueda de un punto de inflexión en la variable de series de tiempo que coincida con la introducción de la política y es el más simple. Si existe confianza que la variable será probablemente afectada por la política, hay un punto de inflexión aparente y en la dirección correcta, se presume que la política ha tenido un efecto; lo abrupto de la inflexión indica el tamaño del efecto. Análisis más sofisticados utilizan

regresiones de series de tiempo u otras técnicas de simulación que definen la situación contrafactual al predecir la tendencia de la variable en ausencia de la política y que comparan la predicción con la tendencia actual (White y Riley, 2002). El éxito de este enfoque depende de la confiabilidad de la tendencia de la variable, de la precisión de las predicciones de regresión y de la estabilidad de las relaciones antes y después de la implementación de la política.

El segundo enfoque depende de la identificación de un grupo ostensiblemente similar al grupo de control, que no esté afectado por la política para ejemplificar la situación contrafactual, y de la comparación de la experiencia de este grupo con la del grupo objetivo de la política. Hasluck *et al.* (2000), por ejemplo, utiliza a las madres con pareja como la situación contraria a la de padres solteros, que son el objetivo de la seguridad social en una política de empleo. El potencial de este enfoque está limitado por la dificultad de encontrar una situación contrafactual y el grado de congruencia entre la situación contraria y la población objetivo de la política.

El tercer enfoque carece, por lo general, de la caracterización de una situación contrafactual explícita y depende de la evaluación de efectividad realizada por los actores de la política (Thornton y Corden, 2002). A ellos se les puede preguntar, utilizando ya sea preguntas abiertas o estructuradas, cuán efectiva consideran la política en términos generales, y tal vez qué tan efectivo juzgan que son sus componentes. Los juicios absolutos de este tipo son algunas veces complementados o reemplazados por medidas relativas que podrían motivar a los entrevistados a comparar el desempeño de la política actual con la reemplazada; lo que es semejante a utilizar a la política anterior como la línea base de percepción o como la situación contrafactual. El valor de tales evaluaciones depende de manera decisiva del conocimiento y del juicio crítico de los entrevistados, pero cuando ocurre un consenso amplio entre ellos, el juicio puede tener una gran confiabilidad (Walter con Williams, 1987). Sin embargo, debido a su naturaleza subjetiva son, usualmente, un elemento de la evaluación retrospectiva de múltiples métodos.

Es importante mencionar, que mientras el seguimiento y las evaluaciones son necesariamente débiles en determinar el costo-efectividad, las evaluaciones de programas han puesto tradicionalmente poca atención en la eficiencia administrativa. Esto no es una consecuencia inevitable de la metodología, sino que es el resultado de las prioridades de investigación. Las evaluaciones de programas a menudo han demostrado satisfactoriamente si una política funciona, pero con menor frecuencia, cómo y porqué.

Es apropiado tener en cuenta las evaluaciones prospectivas y las metaevaluaciones de forma conjunta, a pesar de que la evaluación prospectiva se enfoca al futuro y el meta-análisis extrae sus ideas del pasado. Ambas están presentes durante el ciclo de la política cuando el desafío es considerar qué política podría ser implementada en respuesta a un problema definido —evaluación prospectiva—

y el primer paso es considerar qué políticas funcionaron en el pasado en un contexto semejante —meta-evaluación—.

Meta-evaluación

El meta-análisis es un conjunto de técnicas estadísticas que han sido desarrolladas para agregar y resumir los resultados de los estudios existentes (Lipsey y Wilson, 2001). Desarrollado en la psicología y ampliamente utilizado en la investigación médica, el meta-análisis es comúnmente asociado con la revisión sistemática, que es un conjunto de procedimientos utilizados para identificar el mayor número de estudios relevantes como sea posible (para minimizar el sesgo de selección), y evaluarlos en términos de la calidad de diseño, ejecución y análisis (Glass, 1976). La meta-evaluación formal, el meta-análisis aplicado a las evaluaciones de la política, usualmente comprende un análisis de regresión en el que la variable dependiente es el conjunto de los resultados del programa derivados de un número de experimentos de la política, y el conjunto de variables independientes describen el contenido del programa y el contexto administrativo y ambiental (Ashworth et al., 2002b). El análisis de regresión busca determinar qué características de la política y de su implementación están estadísticamente asociados con las variaciones observadas en los resultados del programa que sirven para determinar qué aspectos de la política y de su implementación funcionan mejor.

Para llevar a cabo una meta-evaluación, la información sobre los resultados del programa, incluyendo el tamaño de su efecto y los errores estándar asociados, es extraída de los informes de un conjunto exhaustivo de evaluaciones relevantes de la política a la par de detalles del diseño del programa, implementación y de los ambientes locales de la política. En las regresiones, los estimados del tamaño del efecto del programa (que es una variable dependiente) se miden por su error estándar para tomar en cuenta con qué diferentes grados de precisión se obtuvieron los estimados del tamaño del efecto.

La meta-evaluación hace más que resumir la evidencia existente sobre lo que funciona, penetra en el funcionamiento de las políticas. Mientras que un solo análisis de programa (o impacto) determinará el costo-efectividad de una política concreta; el análisis formativo asociado (proceso) apuntará hacia los aspectos de su entrega que pudieran contribuir a su eficacia. El meta-análisis puede proporcionar un estimado cuantitativo del valor agregado de cada característica de una política y de su administración, tanto aisladamente como en conjunto. A diferencia de una sola evaluación, el meta-análisis puede además establecer si una política concreta podría funcionar también en una ubicación diferente o si, en efecto, el aparente éxito en la implementación de la política se debe en gran medida a la naturaleza idiosincrásica de su ambiente.

Estudio de Caso 3

Evaluación retrospectiva: evaluación del Fondo Social-Huby y Dix, 1992

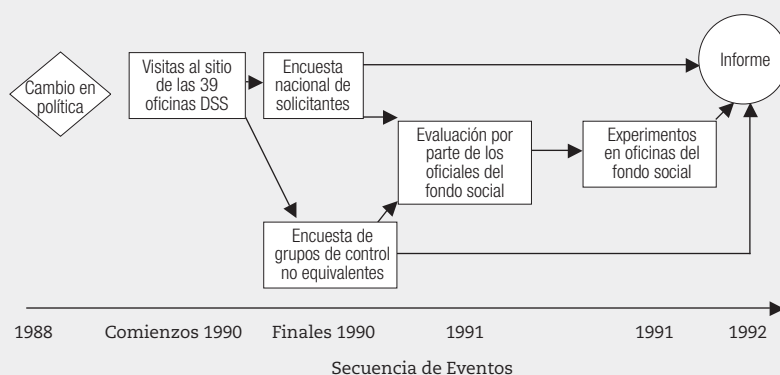
El Fondo Social iniciado en Gran Bretaña en 1997 es un sistema, en gran medida, de préstamos discrecionales y subvenciones que busca satisfacer las necesidades únicas de las personas que viven de la asistencia social, a quienes se les paga de un presupuesto fijo asignado a las oficinas locales.

La evaluación buscaba determinar el alcance al objetivo del Fondo Social enfocado en personas que estaban en una mayor necesidad, un objetivo que se complicó debido a la naturaleza discrecional del presupuesto del esquema. Solamente los oficiales del fondo social podían determinar a su discreción la elegibilidad de las personas, sin embargo, aun su discrecionalidad tenía trabas derivadas de las restricciones presupuestarias.

El enfoque adoptado, después de un proceso de licitación, definió cuatro grupos de comparación no equivalentes (Figura 3.1):

1. Solicitantes exitosos al Fondo Social;
2. solicitantes a quienes se les había denegado un subsidio;
3. beneficiarios del programa de Apoyo al Ingreso que no habían presentado solicitud al Fondo Social para determinar qué tan extensa era la necesidad no satisfecha; y,
4. receptores de Beneficios para la Vivienda que no tenían derecho a hacer solicitudes al Fondo Social, pero que fueron incluidos como los representantes de las familias de bajos ingresos.

Figura 3.1 Evaluando la perspectiva del fondo social y la meta-evaluación



Fuente: OCDE.

El valor de la meta-evaluación variará de acuerdo al número y calidad de las evaluaciones relevantes disponibles y al grado de correspondencia entre el problema de la actual política y aquellos que fueron abordados por políticas previas que han sido evaluadas. Actualmente, la limitación anterior es muy grave dado que el número de evaluaciones de política realizadas en cualquier esfera concreta de la política, y especialmente dentro de la misma jurisdicción, sea probablemente pequeña. Sin embargo, este problema debería reducirse con el tiempo, en tanto más implementaciones de política son sistemáticamente evaluadas, aunque siempre habrá cuestiones acerca de lo comparativo de las evaluaciones y las variables utilizadas en ellas que pueden ser sólo en parte abordadas por demandas a una mejor práctica en el diseño y en el reporte de las evaluaciones.

Evaluación prospectiva-microsimulación

En principio, la meta-evaluación puede ser usada a futuro más que hacia el pasado para encarar la pregunta “¿Qué política funcionaría?”. Una vez que un modelo de regresión ha sido desarrollado, las características de varias opciones de política y su contexto pueden sustituirse en la ecuación para establecer su resultado posible y su costo-efectividad relativo. Sin embargo, las limitaciones en los datos implican que esta forma de simulación de la política tiene aún que tener el mismo impacto que aquellos basados en modelos de microsimulación estática, y más recientemente, de micro-simulación dinámica. Un enfoque que va más lejos de la evaluación prospectiva y que jugará probablemente un papel creciente es la experimentación de laboratorio.

Los modelos de microsimulación pretenden predecir el probable impacto de un cambio de política en personas individuales o en familias, al sumar los efectos individuales que generen estimados del impacto total (Klevmarken, 1997; Plum, 1998). Algunos de los modelos más exitosos y simples han sido utilizados para diagnosticar el impacto de cambios en los impuestos y en los beneficios. Antes del surgimiento de los modelos, el impacto de tales cambios hubiera sido discutido en términos del impacto en una familia “promedio”. Ya que muchas familias no se asemejaban mucho al promedio, tal enfoque era a menudo falso. Construir un modelo del impacto en cada uno en una muestra representativa de familias sobre la base de su ingreso y las circunstancias familiares, tal como fueron abordadas en las entrevistas de la encuesta (o cuando fuera posible utilizando detalles de los archivos administrativos), ha probado ser un método mucho más exitoso. En esencia, los datos de las encuestas sirven como un modelo de la población objetivo, proporcionando para cada individuo un conjunto de atributos endógenos y exógenos, siendo los atributos endógenos susceptibles al cambio generado debido a la nueva política. Este modelo inicial sirve para describir la situación contrafactual, el producto de políticas existentes.

Estudio de Caso 4

Meta-evaluación: descubriendo qué es lo que funciona mejor en la provisión de servicios de seguridad social - Ashworth *et al.*, 2002

Este estudio reunió una base de datos que abarca materiales recogidos de las evaluaciones obligatorias de los 24 programas de seguridad social para el empleo existentes en los Estados Unidos, implementados entre 1982 y 1996, que habían sido asignados de manera aleatoria. La base de datos contiene información que incluye estimaciones de impacto y las características de los 64 programas de políticas extraídas de reportes publicados, complementadas con datos referentes a las condiciones del mercado laboral averiguados en otras fuentes.

Los objetivos del meta-análisis eran identificar cuáles características de los programas eran más importantes para explicar las variaciones en el éxito de los programas, de los que incrementaron los ingresos y redujeron la entrega de beneficios de seguridad social; establecer por cuánto tiempo duraban los impactos; y averiguar si los resultados del programa eran afectados de alguna manera por las diferencias en las condiciones del mercado laboral, o en la composición del número de casos de seguridad social.

El análisis empleó un enfoque de regresión de mínimos cuadrados ponderados, utilizando el interés del impacto del programa como la variable dependiente. La ponderación ajustó la variable del impacto del programa a la varianza de cada una de las evaluaciones originales. La ponderación fue calculada como el error estándar de los estimadores de impacto, de tal forma que los impactos con las menores varianzas recibieron un peso mayor que reflejara su solidez.

El análisis confirmó la visión ampliamente aceptada de que las intervenciones cuyo énfasis está en medidas enfocadas a alentar a las personas a que regresen rápidamente al trabajo, en lugar de comprometerse en la adquisición de capital humano, son las que más reducen el número de los enlistados en seguridad social y dan como resultado mayores ganancias en el ingreso. También indicó que las diferencias estadísticas existentes entre los grupos de acción originales y los de control atribuibles al impacto de las políticas duraban varios años. Sin embargo, el desempeño de las intervenciones de la política era sensible a las diferencias en el medio ambiente local y a las variaciones en las características del número de casos de la seguridad social, y estos factores contribuyeron a explicar el aparente éxito notable en ciertos programas.

Los atributos endógenos son entonces “actualizados”, de acuerdo a un conjunto de probabilidades transitorias que simulan el efecto del cambio de política, que es después “ampliado” de acuerdo a fracciones de la muestra para generar los estimados de la población. Los primeros modelos eran estáticos, en

tanto, solamente proporcionaban los efectos inmediatos de la primera ronda de los cambios de política pero, al incorporar supuestos sobre la conducta en los modelos de microsimulación, fue posible tomar en cuenta los efectos de la segunda y tercera rondas. Así, por ejemplo, los modelos de impuestos y beneficios permitirán cambios en los patrones de gasto, resultantes de las modificaciones en los impuestos indirectos, y cambios en el empleo provocados por modificaciones en la estructura de incentivos debidas a reformas en los niveles de beneficios o en las tasas de impuestos.

El éxito de la microsimulación depende de la calidad de los datos y del modelo. El modelo es, en cambio, dependiente de un entendimiento previo del sistema que se está modelando, así como de la pertinencia y especificidad de la teoría subyacente. Es de importancia crítica, la validez de los supuestos sobre la conducta implícitos en el modelo; éstos se derivan típicamente de una juiciosa mezcla de teorías, a menudo pobremente desarrolladas, y de la evidencia empírica. Además, entre mayor sea el cambio de política, lo más probable será que los supuestos sobre la conducta no sean relevantes en el futuro, especialmente si el modelo ha sido graduado en referencia a los cambios en la conducta observados en un ambiente de política radicalmente diferente. La experiencia sugiere que el modelo puede beneficiarse de la discusión abierta de los métodos y supuestos y aun de la competencia entre diferentes equipos de quienes construyen modelos (Cabinet Office, 2000).

Desarrollar un modelo en un ambiente complejo de política es un procedimiento largo y asociado con altos costos, pero esto puede contrarrestarse por la longevidad de los modelos resultantes.

Evaluación prospectiva — experimentación

A diferencia de la simulación de políticas, la experimentación de laboratorio ofrece la segunda, y en términos generales, ruta complementaria para la evaluación prospectiva² que puede implementarse rápidamente a un costo comparativamente bajo. La experimentación de laboratorio toma una de dos formas básicas pero, tiene un alcance mayor si ambas se mezclan (Huby y Dix, 1992). La primera forma, la *experimentación sobre la conducta* adopta un enfoque positivista buscando medir el comportamiento individual bajo un escenario controlado. La segunda, aquí denominada *evaluación de laboratorio*, crea un escenario artificial en el cual los encuestados son estimulados a reflexionar sobre las opciones de política.

La experimentación de la conducta se utiliza típicamente para establecer los principios de la política más que para evaluar una opción rigurosamente definida de política. Una preocupación clave en el diseño de la política es la probable

Estudio de Caso 5

Microsimulación prospectiva: una evaluación de los nuevos créditos fiscales - Brewer et al., 2001

El gobierno del Reino Unido estaba por implementar en 2003 importantes reformas a los sistemas de impuestos y beneficios con la introducción de un Crédito Fiscal a la Niñez, que reúne la mayor parte del apoyo financiero disponible para los niños en un solo esquema de ingreso, y el Crédito Fiscal al Empleo, que proporciona asistencia financiera a los adultos empleados con empleos poco remunerados.

A diferencia de otros estudios de caso, la evaluación prospectiva no fue financiada directamente por el gobierno, aunque éste les proporcionó asistencia técnica a los investigadores. El objetivo de la investigación era establecer con base en detalles previamente hechos públicos sobre las reformas propuestas:

- el número de probables beneficiarios;
- el impacto en las tasas de pobreza para diferentes tipos de familia; y,
- el efecto en los incentivos al trabajo causado por cambios en la restricciones presupuestarias (resultado del intercambio entre los salarios recibidos por tomar un empleo adicional y la consecuente pérdida en la asistencia financiera como resultado de un mayor ingreso).

El análisis se realizó utilizando TAXBEN,¹ un modelo de microsimulación dinámico del sistema británico de impuestos y beneficios desarrollado por el Instituto de Estudios Fiscales. En lugar de simplemente calcular los efectos financieros inmediatos de las reformas propuestas al impuesto, TAXBEN integra, en esencia, modelos estáticos del sistema de beneficios-impuestos con modelos de comportamiento, a fin de aumentar la comprensión de los “efectos de la segunda ronda” ocasionados por los cambios en la política. Por consiguiente, TAXBEN se integraba a un modelo de oferta laboral para proveer estimaciones sobre los efectos de los cambios en los beneficios o en la tributación directa en los incentivos para trabajar. Esto implicaba calcular las restricciones presupuestarias individuales para cada persona en la muestra y simular el nivel deseado de horas trabajadas bajo diferentes regímenes tributarios.

La evaluación sugirió que 5.7 millones de familias tendrían derecho al nuevo Crédito Fiscal a la Niñez, que las tasas de pobreza entre las familias con niños serían reducidas en tres puntos porcentuales, y que la situación de 1.1 millones de familias con ingresos iniciales moderadamente altos empeoraría debido a las reformas. El análisis de restricciones presupuestarias implícito por las reformas mostró, por ejemplo, que una persona con pareja y sin hijos mejoraría al no trabajar que, si trabajara entre 16 y 20 horas semanales con el salario mínimo nacional.

* TAXBEN: Tax and benefit system (Sistema de impuestos y beneficios).

reacción de la población objetivo; ésta es la respuesta de comportamiento que algunos autores han identificado como el talón de Aquiles de la microsimulación. La experimentación busca establecer esta respuesta directamente, en lugar de deducirla de los resultados observados del comportamiento pasado o de confiar en los informes autoreportados generados por las entrevistas de la encuesta. A los encuestados se les pide que tomen decisiones reales que revelen las preferencias en las cuales están interesados los diseñadores de políticas.

Las técnicas de la experimentación sobre la conducta tienen su origen en la economía experimental, pero con ciertas diferencias cruciales. En la economía experimental, a los encuestados (sujetos de la experimentación) se les presentan opciones en las cuales les resultará caro tener una representación errada de una preferencia, porque se enfrentan a renunciar a dinero verdadero si lo hacen así (Blondel *et al.*, 2000; Eckel y Grossman, 2001). Utilizando las así llamadas preguntas de compensación u opciones, se les plantearía, por ejemplo, a los encuestados que escogieran entre tener diez libras para gastar inmediatamente o quince libras en un periodo de seis meses. Debido a que las circunstancias están controladas, cualquier variación en las preferencias expresadas que pudieran provenir de esta razón se elimina.³ Cuando la experimentación de la conducta se aplica a la evaluación de política, las circunstancias están nuevamente rigurosamente controladas, pero los pequeños riesgos financieros son reemplazados por grandes, que se pueden aproximar al costo de una determinada opción de política. Los encuestados escogen entre conjuntos de opciones diseñados para establecer la estructura de sus preferencias simulando el diseño de la política, sabiendo que una de sus opciones, seleccionada aleatoriamente por los investigadores, será recompensada con el pago en efectivo o por medio de un vale equivalente a la relevancia de lo que está en juego. Por otra parte, en lugar de utilizar típicamente a estudiantes y buscando definir relaciones promedio, se escogen a los entrevistados de la población objetivo de la política con la atención enfocada frecuentemente, en las variaciones de los patrones de preferencias reveladas.

El arte de la experimentación de la conducta es el de inventar opciones compensadas, que dan fundamento a las decisiones clave sobre la estructura de una política diseñada a influir en el comportamiento. Mientras más cercana sea la correspondencia entre las opciones compensadas y las características de la política diseñada para desencadenar cambios en el comportamiento, mejor predecirán los resultados experimentales los resultados de la política, especialmente cuando los encuestados son representativos de la población objetivo.

La evaluación de laboratorio es una forma especializada de consulta acerca de la política que vincula a varios grupos de los actores de la política en la evaluación crítica de las opciones de la política. Puede considerarse como una analogía formativa de la experimentación sobre la conducta enfocada en el análisis crítico, más que en el examen empírico de la lógica subyacente al diseño de la política.

Un formato común es el extendido grupo creativo en el que, a través de muestreo u otro método seleccionado, los actores de la política se reúnen uno o dos días y son informados de los objetivos de la política y de las opciones y se buscan sus opiniones, que se intercambian en grupos moderados de varios tamaños y composición o, cuando fuera lo indicado, cuestionados individualmente (Walker, 1985b). El objetivo puede ser variado, entre explotar la experiencia del grupo para llevar a cabo un análisis FODA (fortalezas, oportunidades, debilidades, amenazas), identificando las fortalezas y debilidades potenciales de la política, conjuntamente con sus oportunidades y amenazas; buscar una respuesta consensuada, o definir el grado de apoyo u oposición. Un objetivo alternativo podría ser desarrollar o afinar el diseño de la política de forma directa en el marco de la investigación, quizás de manera iterativa con grupos que se reúnen repetidamente para trabajar sobre aspectos específicos o con diseños modificados que se pasan de un grupo a otro para su evaluación analítica. La técnica Delphi desarrollada por la Corporación Rand a finales de la década de los setenta es una variante en la cual, el análisis de la política y su desarrollo, se realiza a través de un proceso iterativo de consulta (Goodyear-Smith y Farnell 2001).

Estudio de Caso 6

Experimentación prospectiva de laboratorio: fomento a la educación de los adultos - Voyer et al., 2002

El gobierno canadiense desea asegurar que, cuando menos, un millón más de adultos puedan aprovechar las oportunidades de aprendizaje en los próximos cinco años y el Departamento de Recursos Humanos está examinando actualmente varias opciones para satisfacer este objetivo. El Departamento necesitaba, en particular, establecer cuáles de los diferentes programas o incentivos financieros afectarían el comportamiento de la población adulta. Para abordar este problema, el Departamento encomendó realizar un experimento de laboratorio de gran escala.

El experimento fue diseñado, primero, para determinar qué tipos de asistencia gubernamental funcionan mejor para alcanzar el objetivo de la política de aumentar la inversión en capital humano entre adultos con diferentes antecedentes socioeconómicos y segundo, para establecer los obstáculos que impiden que los adultos aprovechen los beneficios educativos.

El proyecto abarcó una encuesta basada en entrevistas representativas con los adultos para averiguar las características demográficas, socioeconómicas, de comportamiento y de actitud; una serie de preguntas de selección individual relativas a la compensación monetaria para captar, experimentalmente, las preferencias reveladas, y una evaluación de los conocimientos para medir la habilidad y la habilidad percibida para aprender.

Estudio de Caso 6 (cont.)

Tabla 3.2 Preferencias en los préstamos estudiantiles

	Elección A Efectivo	Elección B Hasta este monto en préstamos estudiantiles para educación de tiempo completo o parcial.
Decisión 21	\$100	\$100
Decisión 22	\$100	\$200
Decisión 23	\$100	\$300
Decisión 24	\$100	\$400
Decisión 25	\$100	\$500
Decisión 26	\$100	\$600
Decisión 27	\$100	\$700
Decisión 28	\$100	\$800
Decisión 29	\$100	\$900
Decisión 30	\$100	\$1000

Las preguntas de elección se diseñaron para que la alternativa de efectivo para cada elección de educación bajo los diversos tipos de asistencia financiera fuera la misma. Como resultado, fue posible observar a qué nivel de asistencia financiera, los individuos estaban dispuestos a cambiar de un tipo de asistencia a otro. Relacionar los resultados de la encuesta con los datos experimentales permitió el cálculo de la proporción de personas que invertiría en educación, dados los diversos tipos de asistencia, mientras que el análisis de regresión indicó cómo cambiaban las preferencias de acuerdo al estatus socioeconómico, el miedo al fracaso, la aversión a los préstamos y otras características personales.

Fuente: OCDE.

Las preguntas de elección involucraban una propuesta para inscribirse en un curso educativo con apoyos que variaban en tipo y nivel, con la intención de obtener las preferencias por educación cuando era financiada por una donación, un préstamo de acuerdo al ingreso o por ahorros subsidiados. Las preguntas de selección para los préstamos se presentan en la Tabla 3.2. Durante el experimento, los encuestados escogieron entre A y B para cada pregunta y, al finalizar, se seleccionaba una pregunta al azar y a cada encuestado se le proveyó del pago correspondiente con la elección hecha en la pregunta seleccionada. Por ejemplo, un individuo que seleccionó B en la decisión 28 recibiría un préstamo por \$1600 para inscribirse en educación.

El valor de la evaluación de laboratorio depende del conocimiento de los entrevistados y de la habilidad de los moderadores de la investigación para capitalizarla. Podría ser de la máxima utilidad cuando la propuesta de cambios en la política sea la vía dependiente y que la experiencia actual sea probablemente, afín, al menos cuando un cambio en el paradigma de la política se proponga. Aun

así, la técnica será valiosa en el último escenario, proporcionando información sobre la administración del cambio.

Evaluación formativa

El espacio es inadecuado para hacer justicia a la evaluación formativa que posee una rica historia y se caracteriza por una diversidad de enfoques nacidos de diferentes posiciones ontológicas y de respuestas creativas a implementaciones particulares de política (ver Ritchie y Lewis, 2003; Patton, 2002 para críticas). Para cada pregunta de la evaluación acumulativa, existe un conjunto correspondiente de preguntas de evaluación formativa que buscan explicaciones, o para la comprensión del resultado de la política (Tabla 3.1). Estas preguntas se prestan para la investigación cualitativa y así la evaluación formativa se caracteriza por la dependencia, aunque no dependencia exclusiva, de técnicas cualitativas como las entrevistas a profundidad, los grupos focales y los estudios de caso.

El tiempo verbal en el que se formula cada pregunta influye, a su vez, en el diseño evaluativo pero, no con el mismo alcance que en la evaluación acumulativa. Cuando la evaluación es retrospectiva, los evaluadores formativos necesariamente dependerán más de la evidencia documental y secundaria, como es el caso de la evaluación que está siendo llevada a cabo en la cual, tanto la investigación de observación como de acción son posibles; la última permite el involucramiento en tiempo real de los investigadores con los actores de la política para ayudarlos en afinar el diseño e implementación de la política. De manera similar, es probable que la evaluación prospectiva dependa más de técnicas de proyección, simulación y dramatización, tal vez usadas en conjunto, que en las evaluaciones de laboratorio analizadas anteriormente (Walter et al., 1987; Estudio de Caso 7). Sin embargo, el enfoque más amplio de la mayoría de las evaluaciones formativas es plural y orientado a la investigación, reuniendo cualquier información relevante que sea posible obtener y de las percepciones comparativas y contrastantes recogidas de las perspectivas de diferentes actores de la política y fuentes de datos.

Es importante apearse a esta simple formulación porque el análisis cualitativo está fundado en numerosas, a menudo competitivas, tradiciones. Patton (2002, pp. 132-133) enumera dieciséis tradiciones establecidas, que incluyen a la etnografía (que se interesa en la descripción de culturas), constructivismo (que busca explicar cómo las personas construyen la realidad), interracionalismo simbólico (investiga el significado individual y colectivo concedido a las interacciones entre las personas) y el análisis narrativo (que se centra en los relatos de las experiencias de vida de las personas para entender mejor sus creencias y acciones). Cada tradición difiere en la respuesta dada a las siguientes preguntas profundas:

- ¿Qué creemos sobre la naturaleza de la realidad?
- ¿Cómo sabemos lo que sabemos?

Estudio de Caso 7

Evaluación prospectiva de laboratorio: discusión sobre la Reforma a los Subsidios a la Vivienda - Walker et al., 1987

La implementación en 1982/3 del esquema de subsidios a la vivienda por las autoridades locales del Reino Unido resultó ser extremadamente problemática, lo cual llevó al gobierno de Margaret Thatcher a encargar una revisión independiente del subsidio a la vivienda. Esto, a cambio, proporcionó el contenido a los documentos GreenyWhite sobre las reformas a la seguridad social (Cmnd, 1985a; b).

La Investigación para el Equipo de Revisión ha demostrado que, algunas de las dificultades administrativas eran inherentes a la estructura del esquema, que fue resultado de la fusión de esquemas anteriores dentro de un marco común que había fracasado en gran medida en unificarlos (Walter, 1985b). Esto condujo a una mayor complejidad e hizo que el funcionamiento eficiente del esquema dependiera en tal grado de las relaciones entre agencias que fue difícil mantenerlo. El Documento White abordó muchas de estas cuestiones, pero dejó que ciertos aspectos clave de la reforma fueran objeto de decisión. El Departamento decidió usar la investigación para aprovechar el conocimiento de aquellos directamente responsables de administrar los Subsidios a la Vivienda entre las autoridades locales, para así asegurar que los nuevos procedimientos funcionarían.

La investigación fue guiada por una sola propuesta negociada. Los responsables de la política de subsidios a la vivienda, de una muestra estratificada al azar de 66 autoridades locales, participaron en una serie de talleres de investigación de un día de duración, durante los cuales, en diversas ocasiones, los participantes trabajaron en grupos que variaban en tamaño, de dos a ocho personas dependiendo del objetivo. Dos grupos de funcionarios se reunieron en dos ocasiones una semana después. Su misión era revisar las propuestas detalladas de gobierno y, en la segunda ocasión, sugerir alternativas. Otros seis grupos se reunieron una sola vez para evaluar el conjunto completo de propuestas. Se les dieron dos cuestionarios de autollenado, uno antes y otro después de los talleres. El primero reunía información sobre los indicadores de desempeño, mientras que, el segundo recababa principalmente una respuesta reflexiva a todas las propuestas de política tal como habían sido afinadas durante el curso de los talleres. Se llevaron a cabo entrevistas telefónicas de seguimiento con la mira de enriquecer las respuestas escritas.

Un aspecto de la consulta se interesaba en el cambio en la definición del ingreso utilizado para diagnosticar el derecho de una persona a pasar de ingreso bruto al neto. Inicialmente el gobierno propuso que, en vez de recopilar la información sobre el ingreso neto actual, se utilizara una fórmula para diagnosticar un aproximado "de concepto" sobre el ingreso neto. Los participantes rechazaron la propuesta gubernamental por ser demasiado compleja para el diagnóstico inicial y rápido y por no ser lo suficientemente sensible para asegurar la equidad. En el transcurso de la investigación, los participantes cambiaron de posición a favor del cálculo del ingreso neto actual, enfoque que el gobierno finalmente adoptó.

- ¿Cómo deberíamos estudiar al mundo?
- ¿Qué vale la pena saber?
- ¿Qué preguntas deberíamos preguntar?
- ¿Cómo deberíamos involucrarnos personalmente en la investigación?

A esta lista, Patton le añade el “pragmatismo”, en el que la elección del método es separado de la epistemología y aparejado con preguntas específicas de la investigación. Walker (2003) argumenta que la posición ontológica de los evaluadores que adoptan esta postura:

“es probable que se corresponda al ‘realismo sutil’ (Hammersley, 1992) aceptando que un mundo diverso y multifacético existe, independientemente del entendimiento subjetivo, pero creyendo que es accesible a través de las interpretaciones de los encuestados. Ellos probablemente se esforzarán por ser neutrales y objetivos en cada etapa del proceso de investigación y de esta manera, se generarán conclusiones que son válidas y confiables, unas que son verdaderas en función de las creencias y entendimiento de sus encuestados y potencialmente generalizables a otros contextos. Buscarán descripciones detalladas sobre las realidades tal y como son entendidas por sus encuestados, pero claramente delimitan entre estas interpretaciones y las suyas en el análisis del proceso de tener sentido de la complejidad a través de la simplificación y la estructuración”.

Es probable que los pragmáticos acepten un rol, tanto para las metodologías cuantitativas como cualitativas, y tomen en cuenta la naturaleza complementaria de la evaluación acumulativa y formativa, y es esta última la que enriquece la comprensión de las causas por las que una política funciona o no funciona bien (Ritchie y Lewis, 2003; Walter, 1985a). Sin embargo, los evaluadores que trabajan, basados en algunas otras tradiciones, consideran a la evaluación acumulativa, especialmente la que se basa en la experimentación, como inherentemente fallida e inferior a la evaluación formativa. Pawson y Tilley (1997) abogan por la “evaluación realista”, arraigada en la filosofía científica realista (Lakatos, 1970; Harré, 1986) y adoptan esta postura.

Aun dentro del marco pragmático, existen muchos modelos de evaluación y enfoques, los principales se listan en la Tabla 3.3, adaptado de Patton (2002) que provee una breve relación de cada uno. Los modelos reflejan los diferentes propósitos de la evaluación dentro de los cuales pueden ser utilizados diversos enfoques. En el modelo clásico orientado a los objetivos, una política se evalúa en relación con los objetivos propuestos por los arquitectos de la política. En tal caso, dependiendo del enfoque escogido, la evaluación formativa busca investigar cómo la política es implementada para entender mejor cómo las interacciones entre varios actores de la política pueden afectar el resultado, o para triangular diferentes

perspectivas sobre el funcionamiento y eficacia de la política. La evaluación carente de metas no prioriza estos resultados, que están directamente ligados a los objetivos de la política pero, explora una amplia gama de resultados intencionados y no intencionados y los procesos que les antecedieron. La evaluación sobre las transacciones hace énfasis en las diferentes perspectivas de todos los actores de la política, los formuladores de la política, los administradores, los empleados a nivel de campo, los clientes, etcétera y puede conllevar una interacción directa entre los evaluadores y algunos o todos de estos grupos con la mira de trabajar en colaboración para mejorar la implementación y los resultados. La evaluación enfocada en la utilización prioriza el entendimiento del proceso de implementación de la política en la creencia que, la forma cómo una política se modela al ser entregada, influencia críticamente su éxito probable. Finalmente, los estudios sobre los conocedores colocan al (los) evaluador (es) en el papel de jueces, tamizando y evaluando la evidencia a la luz del conocimiento profesional y personal.

Volviendo a los enfoques específicos listados en la Tabla 3.3, cuatro merecen especial mención. Primero, los estudios de proceso ejemplifican la evaluación formativa y frecuentemente acompañan a las evaluaciones acumulativas. Buscan elucidar la dinámica interna de cómo operan las políticas. Las interrogantes abordadas por los estudios de proceso incluyen las siguientes:

- ¿Cómo se implementa la política en la práctica?
- ¿Cómo entran los clientes al sistema y qué les ocurre a medida que progresan en el sistema?
- ¿Qué experiencias genera la política para los distintos actores de la política?
- ¿Cómo interactúa el personal, los usuarios o clientes y los recursos?
- ¿Cómo se vinculan las creencias y acciones con los resultados?
- ¿Cómo intervienen los actores en los resultados de la política?
- ¿Cuáles son las fortalezas y debilidades de la política que son identificadas por los actores de la política y cuál es la naturaleza y el grado de consenso con respecto a qué tan bien funciona la política?

Segundo, los enfoques sobre la teoría del cambio (Rogers *et al.*, 2000; Weiss, 1996; Chen, 1990) o teoría de la acción (Patton 2002) merecen especial atención debido a la profunda influencia que está empezando a tener en todos los tipos de evaluación, acumulativa y formativa. La característica que comparten es el objetivo de buscar la percepción de las personas de la secuencia de causalidad que, se piensa, relaciona los insumos de la política, con la implementación y los productos. Difieren en términos de las personas cuya visión se busca, el momento del análisis y el uso que se hace de los resultados. Algunos obtendrán un modelo causal de los diseñadores de política, en los inicios del proceso de evaluación y

Tabla 3.3 **Evaluación formativa: modelos y aplicaciones**

<p>Modelos de evaluación</p> <p>Evaluación orientada a los objetivos Evaluación carente de objetivos Modelos de transacción: evaluación de reacción e iluminativa Estudios de expertos Evaluación enfocada en la utilización</p>
<p>Aplicaciones de evaluación</p> <p>Evaluación de resultados Evaluando resultados individualizados Estudios de proceso Evaluación de la implementación Teorías de cambio/acción; modelos lógicos Valoración de la viabilidad de la evaluación Comparando programas: enfoque en la diversidad Evaluación preventiva Documentando el desarrollo en el tiempo e investigando los cambios en los sistemas</p>
<p>Aplicaciones interactivas y participativas</p> <p>Personalizando y humanizando la evaluación Armonizando los valores del programa y de la evaluación Aplicaciones al desarrollo: investigación activa, aprendizaje activo, práctica reflexiva y organizaciones de aprendizaje Investigación de apreciación Investigación participativa y evaluación: valorando y fortaleciendo la colaboración Apoyando el diálogo democrático y la deliberación Apoyando la democracia a través de la utilización del proceso: ayudando a la ciudadanía a ponderar las evidencias y a pensar en términos de evaluación</p>
<p>Aplicaciones enfocadas en la calidad</p> <p>Entendiendo y revelando la calidad Garantía de calidad</p>
<p>Aplicaciones especiales</p> <p>Medidas que no obstaculizan Consideraciones sobre el nivel de desarrollo: ausencia de probada instrumentación cuantitativa Investigación de confirmación y esclarecimiento: añadiendo profundidad, detalle y significado a los análisis cuantitativos Reconocimiento rápido Recolectando y comunicando historias Monitoreo legislativo y auditoría Aplicaciones futuras: investigación anticipada y análisis prospectivo de política Rompiendo la rutina: generando nuevas percepciones</p>

Fuente: Adaptado de Patton (2002).

diseñarán una metodología para probar o establecer la validez de la teoría en la práctica. Otros buscarán modelos causales en diferentes grupos de actores de la política, una vez que la política ha sido implementada, en la creencia que la falta de congruencia en las percepciones podría ayudar a explicar las fallas de la política. Aun, otros se comprometen con los actores de la política para idear y refinar modelos causales con una visión de desarrollar y mejorar el diseño de la política y la implementación. La mayor contribución de los enfoques de la teoría del cambio es la de llamar la atención hacia la necesidad de hacer explícita y transparente a la caja negra de las políticas. Mientras que podría ser posible establecer si una política es un éxito sin una clara comprensión, o al menos la hipótesis de cómo es que debe funcionar, sin tal conocimiento, es muy difícil afinar la política o derivar lecciones que podría tener aplicabilidad en diferentes áreas.

Estudio de Caso 8

Un prototipo formativo: el nuevo pacto para personas con discapacidad —esquemas innovadores— Hills *et al.*, 2001

En 1997, el gobierno del Reino Unido puso en práctica dos variantes del Nuevo Pacto para Personas con Discapacidad. El primero (Estudio de caso 2) empleó asesores personales, mientras que el otro buscó identificar y examinar diferentes enfoques de cómo ayudar a las personas a incorporarse o permanecer en el trabajo. Después de una licitación, se evaluaron y establecieron 24 diferentes esquemas utilizando en gran medida una metodología cualitativa.

La evaluación de los innovadores esquemas piloto fue encargada por el Departamento de Seguridad Social sobre la base de la licitación. Se emplearon una gama de diferentes recursos metodológicos que incluyeron:

- Una auditoría de cada esquema que involucraba visitas de campo al inicio y conclusión del programa de dos años, el análisis documental y las entrevistas con el personal, los participantes y las organizaciones asociadas;
- Estudio temático de casos alrededor de asuntos tales como el trabajo con los empleados y las relaciones con los socios;
- Estudios de caso basados en esquemas para analizar las relaciones entre los esquemas y organizaciones más amplias y las rutas seguidas por los usuarios a través del servicio;
- Análisis de los datos de seguimiento proporcionados por los esquemas al Departamento de Seguridad Social; y,
- Talleres de diseminación utilizados para validar el aprendizaje mediante la presentación de las conclusiones a los representantes de los esquemas.

Durante el curso de la evaluación, los evaluadores desarrollaron un modelo de “ruta” que les ayudó a trazar un mapa de los diferentes tipos de actividades, en las cuales los esquemas estaban involucrados. Este modelo, a pesar de no ser utilizado como un instrumento por los esquemas, indicó la necesidad tanto de movilizar como de apoyar y de avenir y mediar entre las actividades que se daban tanto con los clientes como con los patronos.

La evaluación fue capaz de describir y comentar sobre las trayectorias seguidas por los clientes, sobre los procedimientos adoptados en desarrollar las asociaciones al trabajar con otras agencias, sobre los estilos y tareas de administración y sobre la viabilidad financiera de los esquemas. También buscó identificar a los tipos de clientes a los cuales se dirigían los esquemas y examinar la naturaleza de las relaciones entre el esquema, los clientes y los empresarios.

Tercero y cuarto, la evaluación formativa es a menudo utilizada en situaciones donde las evaluaciones cuantitativas acumulativas son imposibles o inapropiadas. Patton (2002) nota situaciones del “nivel de desarrollo” en donde

no existen medidas aceptables, válidas y confiables. De manera similar, la investigación cualitativa puede utilizarse para confirmar, calificar, interpretar, iluminar e ilustrar a las evaluaciones acumulativas; cuando esta investigación es comprensible o no explícitamente subordinada a la evaluación acumulativa, tal investigación demanda la etiqueta de “evaluación formativa”.

Regresando a la tipología simple introducida en la Tabla 3.1, éste provee un resumen del argumento hasta este punto. Identifica los diferentes propósitos de la evaluación de las políticas, relacionando éstos a las clases de conocimientos requeridos para formular las políticas. Sugiere que diferentes diseños de evaluación son los apropiados para proporcionar información a varias preguntas de las políticas, “sugiere” porque el diseño de las evaluaciones es un proceso profundamente creativo, en el que pocas veces hay una sola e ideal solución metodológica. Finalmente, apunta a la necesidad de contar con evidencia de evaluación en cada estadio del ciclo del proceso de formulación de las políticas. Lograr esto requiere de un compromiso político sostenido de construir la política basada en evidencia y de una infraestructura capaz de proporcionar información apropiada en el momento más oportuno.

Finalmente, los evaluadores fueron capaces de comentar sobre la eficacia de los esquemas. Los objetivos sobre el número de colocaciones que pronosticaban los esquemas fueron abandonados, en tanto, otros fueron juzgados como irreales y la evaluación se enfocó, en cambio, en el porcentaje de clientes reclutados que se incorporaron en el trabajo y sobre qué tan selectivos fueron los esquemas en su reclutamiento. De los esquemas que fueron juzgados como los más “exitosos”, cinco fueron selectivos en cuanto al criterio de estar listo para el empleo y cuatro no lo fueron, mientras que esquemas comparativamente “no exitosos” no fueron tan selectivos en sus procedimientos de reclutamiento.

Amenazas a la evaluación

La tipología de la Tabla 3.1 también presenta un escenario de caso ideal, en el cual la evaluación de la política podría ser conducida y usada para proporcionar información para las decisiones en cada punto del ciclo de elaboración de la política. En realidad, como muchos autores han notado, la mayor parte de las políticas no son formalmente evaluadas y aun cuando lo fueran, los resultados podrían no ser utilizados nunca o simplemente ser explotados para justificar una decisión anterior (Bulmer 1986; Walter, 2000; Weiss, 1992; Wilensky, 1997). Más que simplemente añadir a esta letanía decepcionante, la Bretaña posterior a 1997 es considerada como un estudio de caso que ilustra algunos de los desafíos a los que se enfrenta una administración que está comprometida con la creación de políticas a partir de las evidencias. Entonces, en la sección final, este estudio de caso se utilizará para reflexionar sobre la forma de asegurar o incrementar

la probabilidad de que la buena evidencia de la evaluación sea producida y eficientemente utilizada.

Como otro apartado valioso, vale la pena reconocer las dificultades metodológicas en establecer si cualquier evaluación en particular ha tenido un efecto en la política (Weiss, 1998). Invariablemente falta una situación contrafactual que demuestre lo que habría podido ocurrir en la ausencia de la evaluación, así que uno se ve forzado generalmente a depender de los reportes (Davies *et al.*, 2000; Greenberg *et al.*, 1999; Cabinet Office, 2000). La naturaleza de la burocracia gubernamental, con funciones especializadas, significa que muy poca gente o nadie está en posición de ver el proceso de la política como un todo y de aislar el impacto de la investigación evaluativa. Muchos de los actores clave también tuvieron razones, tanto como para exagerar como para reducir el papel de la evidencia. Más aún, si la tarea de atribuir un cambio en la política, basado en la evidencia producida en la investigación, es difícil, el problema se multiplica cuando cualquier evidencia producida en la evaluación apoya a la crítica política de la política, de forma tal que no se necesita o no es evidente ningún cambio en la política. Además, los estudios sobre el impacto de la investigación de la política (en oposición al papel de la evidencia de la evaluación *per se*) sugieren que los procesos son a menudo indirectos y efectivos sólo en el largo plazo, filtrándose a través de los corredores de poder como el agua que permea la piedra caliza, cambiando lentamente la comprensión de los temas (Thomas, 1987; Weiss, 1980).

Ambientes de política

Volviendo al Reino Unido como un estudio de caso, desde los últimos veinte años se ha producido un cambio radical a la priorización en la formulación de política (Davies *et al.*, 2000; Walter, 1997). Este desarrollo fue estimulado en los años ochenta por el surgimiento de la así llamada “nueva administración pública” (Hood, 1991; Stewart, 1996) y por la introducción asociada de las doctrinas de los administradores y las técnicas para mejorar el seguimiento, control y evaluación del desempeño (Carter y Creer, 1993). Con los principios establecidos, en 1988 el Tesoro de su Majestad publicó la primera edición de una perdurable guía para la evaluación de políticas para los administradores. Ésta incluía un “marco de evaluación” que llevaba al administrador de la política a través de la discusión del propósito, la naturaleza y la ejecución de la evaluación (Treasury, 1988). En su mayoría, la guía estaba limitada a la evaluación de las políticas existentes, aunque hacía referencia al papel de la evaluación en la valoración de la política.

Estos cambios fueron acompañados por un crecimiento sustancial en la evaluación de políticas durante la década de los ochenta y noventa (Walter, 2000b). Usualmente, esta evaluación era retrospectiva en lugar de prospectiva. Por lo tanto, dependía de enfoques plurales más que de la metodología de asignación

aleatoria, que por ese entonces se estaba convirtiendo en el pilar de la política de evaluación en los Estados Unidos. Las razones para este enfoque distinto fueron principalmente estructurales. Gran Bretaña es altamente centralizada, con muchas políticas que están siendo implementadas de manera uniforme directamente por las agencias del gobierno central, mientras que el sistema federal de los Estados Unidos delega un mayor diseño de las políticas a nivel estatal. Hasta 1996, cambiar las provisiones de la seguridad social y de la asistencia social era ilegal en Gran Bretaña, lo que impidió a los diseñadores de políticas llevar a cabo experimentos de política basados, ya fuera en la asignación aleatoria o en áreas controladas. En contraste, en los Estados Unidos, la experimentación, por supuesto en el área de bienestar, se convirtió en un mecanismo de lucha de los estados por mayor autonomía y recursos. El Acta de Apoyo a las Familias de 1988 permitió a los estados modificar las políticas financiadas federalmente (a través de la concesión de los “no aplicantes”) sólo si la innovación en la política era evaluada por medio de una metodología de asignación aleatoria. Es además probablemente cierto decir que la evaluación de políticas en la Gran Bretaña está menos influenciada por la economía positivista que en los Estados Unidos, y que la sociología y que las ciencias políticas están más influenciadas por el constructivismo y las tradiciones intepretativas. Allí hubo, por lo tanto, menor demanda por la experimentación y una mayor aceptación del valor de la evaluación formativa.

Un compromiso hacia la política basada en evidencias

La elección del gobierno laborista en 1997 empujó la pauta del cambio. Estaba comprometido con la modernización de las políticas y la formulación de políticas basadas en las evidencias y realizar casos piloto antes de la implementación. También abogaba por el incremento en la responsabilidad del gobierno basada en la publicación de los índices de desempeño de la política en contraposición con los objetivos de política preestablecidos (Walter, 2001; Cm., 2002). En el poder, el gobierno laborista publicó el documento White “*Modernizando al Gobierno*” (Cm., 1999) que promovió la formulación de políticas enfocada en los resultados, conjuntamente con un manual normativo para la “*Formulación Profesional de Políticas*” (Cabinet Office, 1999), que enfatizó el uso de evidencia y argumentó a favor de “construir una evaluación sistemática de los primeros resultados en el proceso de la política”. Se estableció una Unidad de Desempeño e Innovación dentro de la Oficina del Gabinete que, a su vez, produjo otro manual que buscaba colocar al buen análisis en el centro de la formulación de las políticas (Cabinet Office, 2000; Davies et al., 2000).

La naturaleza centralizada del Estado británico aseguró que la meta de encargar evidencias para la política pudiera alcanzarse fácilmente. Cerca de 70 políticas piloto — evaluaciones de programa y prototipos— fueron iniciadas por los departamentos del gobierno central entre 1997 y 2002, dejando de lado a las

otras investigaciones, oficialmente encargadas, relacionadas de política.⁴ En su mayor parte, estas evaluaciones fueron llevadas a cabo por organizaciones independientes de investigación, contaron con los fondos suficientes y utilizaron los mejores diseños posibles dadas las restricciones que se discuten adelante. Más aún, los diseñadores de las políticas estaban ansiosos por conocer los resultados de las evaluaciones, que ellos habían encargado, y las conclusiones ciertamente llegaron hasta los escritorios de quienes toman las decisiones. Los resultados de la evaluación también fueron publicados en su mayoría y algunos informes atrajeron el comentario de la prensa. No cabe duda que, en muchas ocasiones, las políticas tomaron forma a la luz de las conclusiones de la evaluación (Davies et al., 2000).

Por tomar un ejemplo, el Nuevo Pacto para los Desempleados de Largo Plazo implementado a nivel nacional en junio de 1998. Esta medida de activación fue menos elaborada y menos flexible que otras introducidas por el gobierno laborista y un temprano seguimiento administrativo demostró que estaba siendo menos efectiva en ayudar a la gente a encontrar empleo (basado en datos brutos que no tomaban en cuenta ni los efectos de sustitución ni los de peso muerto). Para noviembre de 1998, se habían lanzado proyectos piloto en 28 áreas para probar estrategias alternativas y éstas revelaron que reduciendo el tiempo de espera en el acceso al programa, sumando una fase de actividad intensa y alentando mayor flexibilidad en el enfoque, todas juntas habían incrementado las cifras de quienes dejaban los beneficios por el empleo en un 73% (nuevamente, sólo se utilizaron estimados brutos; Lissenburgh, 2001). A la luz de esta evidencia, el esquema nacional fue remodelado. Las personas desempleadas fueron admitidas a los 18 meses, en lugar de los 24, periodos obligatorios de actividad intensa modelados en otros programas fueron agregados y paquetes flexibles de apoyo intenso fueron introducidos, que incluían experiencia laboral y colocaciones, entrenamiento ocupacional y el desarrollo de habilidades menores.

Visibilidad política

Sin embargo, algunas características del ambiente de la política conspiraron en contra, tanto de la efectiva utilización de las evaluaciones de la política, como de los diseños más eficientes. Primero, el sistema político británico es muy confrontador y los resultados de las evaluaciones de las políticas constituyen municiones en el debate político. Muchas de las políticas promovidas por los laboristas, antes de su elección en 1997, fueron políticamente controversiales y los ministros sentían la necesidad de justificar la implementación de las decisiones refiriéndose a la evidencia obtenida de las evaluaciones y de defenderse en contra de los ataques de los grupos de oposición. En octubre 1997 —sólo dos meses después del inicio de la evaluación de dos años del Nuevo Pacto para Padres Solteros (una medida de activación del empleo orientada a los padres solteros que recibían asistencia social)— la Secretaría de Estado para la Seguridad Social

reportaba que uno de cada cuatro padres solteros estaba encontrando trabajo (Dss, 1997). Asimismo, los voceros de la oposición seleccionaron las conclusiones más negativas de los informes de evaluación.

Mientras que, no es nuevo que la investigación sobre evaluación sea utilizada para justificar decisiones *a priori*, o tomadas exclusivamente sobre bases ideológicas (Greenberg *et al.*, 1999; Walter, 1987), el riesgo es que tal lectura selectiva lleve a la investigación evaluativa, conjuntamente con las “estadísticas”, al descrédito público. Una posibilidad adicional es que los ministros decidirán que el agravamiento político generado por la evaluación supera su valor como una herramienta de política.

Sincronización (sentido de la oportunidad)

Segundo, el ritmo en la elaboración de políticas durante la administración laborista de 1997 fue excepcionalmente frenético y los ministros frecuentemente procedían con la total implementación mucho antes de que terminara el programa piloto. Esto había sido previsto para ciertas políticas como el Nuevo Pacto para Personas con Discapacidad (una medida de activación dirigida a quienes reciben beneficios por discapacidad) cuando se esperaba que las decisiones se tomaran a los doce meses en un periodo de evaluación de dos años (Loumidis *et al.*, 2001). Sin embargo, en otros casos, nuevas consideraciones políticas se adelantaron a la espera de resultados. El Nuevo Pacto para Padres Solteros provee un ejemplo de implementación nacional planteada en el contexto de una feroz oposición a las propuestas del gobierno de reducir los diseñadores pagados a los padres solteros (Hales *et al.*, 2000). En tanto que, los primeros resultados son raramente atípicos, los diseñadores de políticas que, sucumben a la tentación de depender de conclusiones provisionales, se arriesgan a cometer errores significativos.

Los calendarios de investigación y de política rara vez coinciden [a pesar de que es más probable que la investigación tenga el mayor impacto si los resultados están disponibles cuando los formuladores de las políticas están buscando información (Berthoud, 1984; Walter, 200b)]. Sin embargo, en Gran Bretaña esta incompatibilidad se acentúa más con respecto a la evaluación prospectiva. La duración de los parlamentarios y el periodo promedio de gestión de dos años de los ministros —un breve periodo durante el cual los políticos necesitan demostrar logros— conspiran conjuntamente en contra de los prolongados experimentos de política comunes en los Estados Unidos.

Muchos de los programas piloto del gobierno laborista fueron prototipos más que evaluaciones de programas, con varias formas de seguimiento temprano que fueron usadas para afinar delicadamente la implementación de la política y, en una menor medida para un diseño de política. Más aún, las decisiones claves de política fueron a menudo tomadas con información de un seguimiento elemental

y con los primeros resultados de las evaluaciones del proceso. En este contexto, no es evidente que la escala de los programas piloto y el conjunto de métodos de evaluación y “controles” utilizados como un intento para asegurar el rigor científico estuviesen garantizados.

El poder de los políticos

Tercero, en Gran Bretaña, los funcionarios civiles están en sus puestos para servir a su ministro (y a través de ellos, al monarca). Esto confiere a los ministros gran autoridad, y cuando están interesados, gran influencia en la forma de las evaluaciones de política, algunas veces hasta en las minucias del diseño. Esto también implica una gran responsabilidad para los funcionarios en su argumentación por casos de diseños sólidos, pero significa que ellos tienen autoridad limitada para insistir y existen numerosos ejemplos, aún desde 1997, de consideraciones políticas o miedos que llevaron a los diseños por debajo de lo óptimo. El rechazo a la asignación aleatoria en muchas ocasiones es un caso en concreto (Walker, 2001).

Complejidad del gobierno

Cuarto, mientras que Gran Bretaña carece de las complejidades de un gobierno federal, los problemas de las políticas generalmente se extienden más allá de la competencia de un solo departamento y muchas agencias están generalmente involucradas en la entrega. Esto significa que el propósito y el método de evaluación tienen que ser concertados entre departamentos, resultando en una investigación compleja y algunas veces sobre elaborada y en una negociación prolongada sobre el diseño de los instrumentos de investigación, etcétera, lo que frecuentemente disminuyen el tiempo disponible para el trabajo de campo y el análisis.

Hiperactividad

Quinto, la amplitud del manifiesto laborista de cambio significó que la mayor parte de las áreas de las políticas estuvieran en un estado de inestabilidad entre 1997 y 2001. Al nivel más básico, el entusiasmo del gobierno en la realización de programas piloto de políticas y controles geográficos rápidamente agotó la oferta de las localidades, que no estaban siendo utilizadas como áreas de acción o de control. Más profundamente, el amplio alcance de los objetivos de política limitó el número disponible de variables independientes de control. La mayor parte de las nuevas políticas de bienestar tenían tanto objetivos macroeconómicos explícitos como de comportamientos individuales. Esto hizo difícil controlar los cambios exógenos, por ejemplo, en la actividad económica, cuando la política

en cuestión debía operar en una jerarquía de niveles (Anderton, Riley y Young, 1999b).

Amnesia de la política

Finalmente, si la naturaleza antagónica del sistema político británico lleva a un cortoplacismo, también promueve la amnesia de las políticas. Es una tradición británica, que los gobiernos entrantes no tengan acceso a los archivos de política creados por sus predecesores y, mientras que los funcionarios civiles pueden proveer alguna forma de memoria institucional, el modo más comúnmente practicado en la formulación de políticas es el de empezar de nuevo desde el principio. Este enfoque incide negativamente en contra de la sistemática acumulación sobre la inteligencia de las políticas y sobre los datos de las evaluaciones, y promueve la aceptación de que la investigación concerniente a las políticas previas es de una relevancia y valor limitados. Mientras que la mayoría de desarrollos de las políticas comienzan con una búsqueda de la evidencia existente, estas búsquedas se asemejan más a viajes de descubrimiento que a solicitudes de información.

La experiencia británica sugiere que, aún con un compromiso hacia la evaluación de políticas emanado desde lo alto del gobierno, el ambiente político no apoya naturalmente la formulación de políticas basada en evidencias. En tanto, la elección del gobierno laborista en 1997 generó una evaluación de las políticas, encargada directamente y publicada por el gobierno, convirtiéndose en una herramienta omnipresente de política, el ambiente de las políticas aún frustra al más efectivo diseño y a la utilización de la evaluación. Alguna de las razones para esto, tal como la amnesia de las políticas, se refuerza por los precedentes documentados en forma local y no generalizada, de tal forma que pudiera sentar jurisprudencia. Otras, incluidas la interferencia política y la incongruencia entre la investigación y los calendarios de las políticas tienen mayor aplicación universal.

Características de la política

La experimentación clásica tiene como premisa la medición del efecto en una variable —el objetivo de la política— al manipular la otra —el instrumento de la política—, mientras que todas las demás variables permanecen constantes. En la realidad, las políticas a menudo tienen una multiplicidad de objetivos que pueden competir, diferir entre los departamentos de gobierno y los ministerios, y cambiar a lo largo del tiempo. Cuando existen múltiples objetivos, y por ende múltiples variables de resultados, los intercambios en la eficacia de los diseños de evaluación son casi siempre inevitables. Como ya se mencionó, las cuestiones de diseño se vuelven aún más complejas cuando algunos objetivos se

relacionan con cambios en la conducta individual y con los efectos agregados. Ya que no todos los resultados pueden ser medidos con igual precisión. Existe una necesidad de priorizar los objetivos de la política, al menos para los propósitos de la evaluación; esto es algo que los políticos no siempre desean hacer y que a veces produce como resultado, diseños ineficientes.

Este problema es, por supuesto, exacerbado cuando es improbable que el impacto de las reformas en la política o en cualquiera de los objetivos sea grande. Más aún, es probable que ésta sea la norma más que la excepción. White (2000), por ejemplo, examinando los esquemas de los programas laborales de seguridad social a lo largo de Europa encontró que: “el impacto de la mayoría de los programas caía dentro de más o menos cinco puntos de una ganancia de diez puntos porcentuales en el empleo” y de manera similar, efectos modestos han sido reportados por la vasta mayoría de evaluaciones hechas aleatoriamente sobre las reformas a la seguridad social en los Estados Unidos (Ashworth et al., 2002b). Si todo lo que puede esperarse son pequeños efectos, esto tiene implicaciones importantes para el tamaño y estilo de las evaluaciones. Puede significar muestras muy grandes en tamaño, algunas veces llevar a cabo encuestas basadas en entrevistas que son prohibitivamente caras o requerir una mayor o total dependencia de las estadísticas administrativas. En Gran Bretaña, los archivos administrativos generalmente carecen de información importante acerca del contexto —ya que la ley solamente permite recolectar y retener información concerniente a la determinación del derecho de una persona a beneficiarse de una política—, por lo que se requeriría de una legislación que permitiera que la calidad y la relevancia de tal información fuese resaltada. Un enfoque alternativo sería aumentar la precisión de las evaluaciones de la política mediante el diseño, más que basándose en la escala, pero, las oportunidades son limitadas siempre que los objetivos múltiples de la política son importantes.

El diseño eficiente no sólo requiere priorizar los objetivos, también necesita especificaciones sobre el grado en que el cambio que constituiría un éxito de la política. Esto raramente se ha hecho en Gran Bretaña.⁵ Si la magnitud de la intensidad de cambio no se especifica a priori, es posible que después del evento se declare que casi cualquier cambio observado en la dirección preferida sea un éxito. Ciertamente, la diseminación de los resultados de las evaluaciones de la política en Gran Bretaña han estado sujetos a un giro positivo, aunque los informes publicados sobre las evaluaciones son detallados y parecen ser escrupulosamente exactos.

En Gran Bretaña, los diseñadores de política, que elaboran programas piloto, son propensos a experimentar con una gama de permutaciones de la política. Cada permutación agregada a un diseño experimental puede, tanto reducir el tamaño efectivo de la muestra, como requerir de una muestra más grande, y por lo tanto, de un programa piloto más caro. Este problema es exacerbado por el deseo comprensible de probar las políticas en ambientes contrastantes, lo que

significa que cada variante necesita implementarse en cada clase de localidad, en una escala que siendo lo suficientemente grande permita determinar la efectividad de las permutaciones de la política. La evaluación de la Asignación para el Mantenimiento de la Educación, un subsidio para alentar a las personas jóvenes a permanecer en la escuela, se inició con cuatro variaciones de la política en nueve áreas, a las cuales se agregaron posteriormente nuevas variaciones (Ashworth et al., 2002a). En realidad, a pesar de todo, es muy difícil determinar si las variaciones en los resultados observados entre las áreas se deben al contexto local o al proceso de implementación. Los estudios sobre el proceso han ayudado, pero gran parte de las variaciones encontradas entre áreas en las evaluaciones británicas quedaron sin explicación (Loumidis et al., 2001).

Algunas de las políticas evaluadas en Gran Bretaña se especificaron previamente de manera deficiente, en parte, debido al compromiso de darle forma a las políticas de acuerdo al conocimiento disponible en el campo. Muchas de las políticas de activación del mercado laboral del “Nuevo Pacto” han transferido parte de la formulación detallada de la política, al nivel de los asesores personales, quienes se relacionan con los buscadores de empleo o con las organizaciones sociales involucradas en la implementación. En tales circunstancias, ha sido difícil determinar qué o cuál política estaba siendo evaluada, debido que se consideraba al paquete como un todo, lo que no dejó claro a los diseñadores de políticas qué componentes de la política eran importantes para generar resultados y cuáles no lo eran.

Los diseñadores de políticas también han buscado modificar la política a la luz de los resultados de las fases tempranas de la implementación piloto, un deseo comprensible, pero muy difícil de acomodar para los evaluadores del programa. En la práctica, por supuesto, los cambios pragmáticos a la política se realizan a menudo durante la evaluación, ya que es irracional proceder con un esquema ineficaz. La evaluación acumulativa requiere que una política uniforme se implemente uniformemente a lo largo del periodo bajo estudio: si ocurre un cambio que mejora la eficacia, el verdadero impacto de la política se subestimarán, pero si el cambio no tiene efecto, esto no se puede detectar formalmente. Los cambios en la política pueden, por lo tanto, integrarse en general en las evaluaciones acumulativas reiniciando la evaluación o alargando el periodo de monitoreo para generar suficientes casos y comparar los resultados antes y después de las modificaciones. En Gran Bretaña, la evaluación del Nuevo Pacto para Personas con Discapacidad fue reiniciado efectivamente una vez se reconoció que los asesores personales no habían priorizado suficientemente los resultados relacionados con el empleo (Loumidis et al., 2001). Los estudios de proceso pueden ayudar a resolver este tipo de enigmas, al ofrecer conocimientos sobre cómo podría haberse efectuado el cambio.

Para resumir, el intento de construir una política de evaluación como el centro de la formulación práctica de políticas en Gran Bretaña ha requerido, desde

1997, de la evaluación de políticas complejas diseñadas de manera deficiente e inestable. En efecto, las realidades de la política han sentado precedente sobre los requerimientos para una evaluación efectiva. Sin los beneficios del poder de la asignación aleatoria, de metas claras de la política o de una implementación estable, las estimaciones sobre la magnitud del efecto no han sido sólidas y las conclusiones de los impactos de la política han sido generalmente equívocos. Esto ha resultado en frustración, tanto para los diseñadores de política, a quienes se les requiere que usen el valor del dinero en sus argumentos en las negociaciones de recursos y para quienes están encargados con la tarea de esbozar diseños y regulaciones detallados de política (Walker, 2001). Esto ha llevado a una mayor dependencia en la evaluación formativa con diseños metodológicos mixtos, ya que esto usualmente genera evidencia aun ante la ausencia de efectos medibles. [Existen informes de los Estados Unidos que los funcionarios estatales conceden mayor énfasis a los elementos del proceso que a las estimaciones del impacto, cuando deciden sobre opciones de la política, Greenberg *et al.* (2000).

Comisionando evaluaciones⁶

El proceso de obtención de las investigaciones en Gran Bretaña configura, considerablemente, el diseño de las evaluaciones de la política. Los contratos son comúnmente asignados, de manera competitiva a organizaciones seleccionadas, que han sido invitadas a responder a diversas especificaciones detalladas sobre la investigación en un periodo de cerca de cuatro semanas. Los contratistas potenciales son alentados generalmente a reunirse con los diseñadores de las políticas por departamentos y con los investigadores para afinar su percepción de los requerimientos de la evaluación antes de someter una propuesta. Las propuestas recibidas son revisadas, a menudo enviadas a árbitros externos, y a una o más organizaciones son invitadas a presentar sus propuestas oralmente. Usualmente, las negociaciones continúan luego de presentadas a las propuestas, algunas veces involucrando a más de uno de los licitantes para revisar diseños (algunas veces adoptando ideas propuestas por contratistas que no tuvieron éxito) y para fijar el programa final de trabajo y el precio del contrato.

Las especificaciones de la investigación expedidas a los contratistas potenciales son a menudo muy detalladas. Esto refleja las aspiraciones del gobierno de encargar un producto que esté confeccionado según sus necesidades, pero sirve de forma significativa para limitar el alcance de un diseño creativo. Típicamente, por ejemplo, las áreas piloto y de control habrán sido ya seleccionadas bajo una mezcla de criterios basados en la investigación, pragmáticos o políticos, antes que se produzcan las ofertas. El resultado es la fijación no sólo de la localización de las evaluaciones sino también del modelo de evaluación (por ejemplo, cuasiexperimentación con controles basados en

áreas). El calendario también habrá derivado de las prioridades políticas y de la implementación, más que por sus consideraciones de diseño.

La invitación a ofertar típicamente induce a diseños alternativos pero, esto puede ser una estrategia de alto riesgo, cuando se ve desde la perspectiva del contratista potencial. Cuando más de una agencia gubernamental se involucra en el diseño de las especificaciones, que probablemente impliquen negociaciones extensas, los contratistas podrían concluir correctamente que es baja la probabilidad de que todas las agencias acepten una desviación radical del diseño del núcleo de la investigación. De hecho, un arte de los oferentes ganadores es el de ponerse adelante de la competencia, adular al personal del departamento y apegarse al diseño original.

Asimismo, en las reglas del juego, todas las partes reconocen que el horizonte temporal permitido para realizar la evaluación es reducido y que sólo pueden ser valorados los efectos más inmediatos de las innovaciones de la política. Los contratistas de las evaluaciones generalmente señalan esta deficiencia pero, ansiosos de comprometerse en un trabajo interesante y a veces lucrativo, asumen la visión de que cualquier evaluación es mejor que ninguna.

Frecuentemente la realidad es aún peor que la teoría. El imperativo de entregar resultados a tiempo, que proporcionen información para tomar las decisiones de política, a menudo significan que las demoras en la aprobación de los instrumentos de investigación o sobre el trabajo de campo no se reflejan en el calendario de informes. El resultado es que el análisis se contrae, arriesgándose a cometer errores y no se garantiza el contar con una investigación completa de todos los temas. Es común tener que analizar en cuatro u ocho semanas los resultados de una entrevista de 50 minutos con una compleja muestra nacional.

Vale la pena notar que muy pocas organizaciones de investigación tienen la capacidad de llevar a cabo evaluaciones a gran escala sin apoyo adicional, especialmente si se espera que el trabajo se inicie tan pronto como se otorga el contrato (y a menudo antes de que sea firmado). Esto significa que una parte importante del corto periodo destinado a ofertar, se invierte en formar consorcios y en determinar la administración interna y la estructura financiera. De hecho, compilar una oferta competitiva puede fácilmente tomarle días a 20 personas, este tiempo a menudo comprime a una o dos semanas del periodo de ofertas —lo que no necesariamente es la forma ideal de concebir un diseño efectivo de evaluación—. Generalmente, la velocidad en el proceso de ofertar también sirve para dejar fuera a la mayoría de académicos cuyos otros compromisos les impiden responder con la suficiente rapidez.

El desarrollo del diseño de las evaluaciones está imbuido dentro del proceso de la política y está afectado por todas las restricciones anotadas anteriormente. A diferencia de los proyectos de demostración en los Estados Unidos, donde la

importancia de desarrollar un diseño de evaluación sólido generalmente antecede a casi todas las demás consideraciones, las cuestiones sobre diseño ocupan un segundo lugar detrás de las preocupaciones políticas que generalmente son tratadas al final. De ahí que los investigadores —tanto quienes están dentro de la maquinaria gubernamental, como los contratistas potenciales— se enfrenten a la tarea de crear una evaluación para la implementación de un programa piloto en especial y que no se les pregunte cuál es la mejor forma de evaluar una política. La propuesta que contiene nuevas evaluaciones de las políticas diseñadas para promover la retención y la promoción en el empleo es una importante excepción y pretende ser el primer proyecto de demostración en Gran Bretaña. El mismo fue desarrollado deliberadamente desde la Oficina del Gabinete, externa al principal departamento de políticas.

Este último desarrollo es ilustrativo de que tanto la naturaleza del proceso de las políticas, como la rigurosidad de la competitividad de las ofertas son perjudiciales para la evaluación efectiva. Como respuesta, la Asociación de Investigación Social elaboró, actualmente (un cuerpo profesional que representa la comunidad de investigación) un código de prácticas relacionadas con la asignación de contratos de investigación donde involucra activamente a los oficiales gubernamentales. Sin embargo, la realidad política puede fácilmente descarrilar la mejor de las intenciones.

El conocimiento del mercado

El gobierno centralizado de Gran Bretaña crea un comprador monopólico en el mercado de las evaluaciones de política. Esto reduce el número y, potencialmente, la independencia de los proveedores. Podría también contribuir a perpetuar el debate mal informado sobre las políticas públicas que se lleva a cabo en la prensa y en otros medios.

En general, los gobiernos locales carecen de recursos para encargar evaluaciones a gran escala de las políticas. Además, a diferencia de los Estados Unidos, el reducido número de fondos de inversión para la investigación en Gran Bretaña normalmente evita que el gobierno local se involucre directamente en las evaluaciones de política, y que esté más dispuesto a invertir en investigaciones que exploren la naturaleza de los problemas sociales; existe quizás la expectativa de que sea el gobierno central el que debe desarrollar y probar las opciones de política. Como resultado de ello, hasta recientemente, la demanda a gran escala de evaluaciones de política estaba restringida y el número de proveedores, el nivel y profundidad de experiencia eran, en consecuencia, limitados. Desde 1997, el gobierno central ha buscado el conocimiento existente en los Estados Unidos para llenar los vacíos en la capacidad.

Por numerosas razones, la participación de los académicos en las evaluaciones de política ha sido limitada. Como un ministro admitió en el 2000, “desde hace tiempo, existe una oleada de anti-intelectualismo que se expande a través del gobierno, tanto a nivel político como entre los funcionarios” que ha servido “para alienar a la academia” (Blunket, 2000, p. 16). Ciertamente, es poco el uso explícito de la teoría de las ciencias sociales en los estudios británicos de evaluación, ya que los diseñadores de políticas se inclinan a considerar la teoría como jergonza y prefieren tratar con “hechos”. En respuesta, la academia británica ha considerado a la investigación aplicada sobre las políticas gubernamentales con menor estima que para la investigación y a la investigación sobre las políticas gubernamentales con mucha menor estima aun. Falta de entusiasmo para dedicarse a las investigaciones gubernamentales podría reflejar la preocupación sobre el modo en el que se restringe a los académicos por las consideraciones políticas y de la política pública discutidas anteriormente, y por el temor a la interferencia del gobierno en la diseminación de los resultados. Adicionalmente, una parte importante de la comunidad británica de las ciencias sociales ha reaccionado en contra, tanto del enfoque de investigación positivista como del empírico, y hasta muy recientemente, la formación en métodos de investigación, en especial de los cuantitativos, ha sido muy pobre en términos de los estándares internacionales. Asimismo, el proceso competitivo de licitación, con un rápido tiempo de respuesta, ha tendido a favorecer a los investigadores de tiempo completo, que trabajan en unidades especializadas de investigación aplicada, sobre los académicos convencionales.

Si la demanda por las investigaciones basadas en la evidencia continúa, existe un peligro real de ampliar la brecha entre la ciencia social aplicada y la teórica. Por esto, es posible que la floreciente industria de evaluación adquiera más de las características de la investigación de mercado, en tanto que la conexión con la academia se restringe al reclutamiento del personal más joven de los institutos de investigación aplicada.

La naturaleza del monopsonio existente en el mercado de evaluación británico podría entonces influenciar la naturaleza del debate político acerca de los resultados de los estudios de evaluación. Existen varias audiencias potenciales para las evaluaciones de política además de los diseñadores de políticas. Éstas incluyen no sólo a los políticos de la oposición, sino a otros grupos de interés y por supuesto, a los administradores y personal de las agencias gubernamentales y de las organizaciones involucradas en la implementación de las políticas. La comunidad académica y de investigación también debería estar interesada ya que, a través de los medios de comunicación, el electorado indirectamente financia las evaluaciones y se presenta como el mayor beneficiario si la investigación está bien realizada. Es probable que cada audiencia tenga diferentes preocupaciones, y que esté interesada en diversos aspectos de la evaluación, lo cual requiere que

la información sea difundida de diversas formas. Sin embargo, la naturaleza prescriptiva de cómo se encargan las investigaciones incluye el modo y el tiempo, o incluso el contenido del informe, y se elabora de acuerdo a las necesidades de los diseñadores de políticas del gobierno central. Mientras que ahora ya se puede esperar que se publique formalmente, excepto en aquellos raros casos donde se estima que la investigación no ha alcanzado un estándar mínimo, usualmente ésta se produce varios meses después de que se entrega el informe escrito y que se discute con los formuladores de las políticas. Sin embargo, los informes tienden a ser largos y no fácilmente asequibles a una audiencia no profesional; generalmente se emiten comunicados de prensa pero, la cobertura de los medios es, por lo general, limitada. Sólo en algunas ocasiones, diferentes versiones de un informe son destinadas a diferentes audiencias y sin una promoción, estos informes se convierten en documentos de archivo antes que en agentes de cambio.

Finalmente, es importante reiterar la falta de aprendizaje acumulativo en Gran Bretaña. Los conjuntos de datos generados por las evaluaciones de las políticas generalmente han sido archivados (después de volverlos anónimos) con la garantía del acceso público, pero comparativamente pocos han sido sujetos a un análisis secundario. Esto puede deberse simplemente a la ignorancia o a la falta de fondos. Más, probablemente, refleja una falta de interés académico en los temas de las políticas y en el fracaso de la comunidad de las políticas en apreciar la importancia de la investigación básica y del desarrollo teórico para analizar los temas de las políticas y concebir respuestas eficaces. Dado el tiempo limitado durante el cual es, permitido el análisis primario, esto constituye una gran subutilización de recursos.

Para resumir, Gran Bretaña, especialmente desde 1997, ha buscado integrar a la evidencia y a la evaluación de políticas al proceso de formulación de las políticas. A nivel operativo, en gran medida, ha tenido éxito: los diseñadores de las políticas confían en que las mismas serán evaluadas, y en muchos casos, lo son. Sin embargo, que las prácticas actuales sean sostenibles queda sujeto a serias dudas. Los resultados a la fecha, especialmente en términos de establecer el impacto y el costo-efectividad han sido desalentadores y las estructuras y la formulación de las políticas se han acomodado poco a los requerimientos de una evaluación efectiva. Un único modelo de evaluación —concebido libremente como un prototipo— domina, a expensas de otras estrategias que podrían acomodarse mejor a algunos de los requerimientos de las políticas. El pensamiento estratégico, tanto con respecto al desarrollo de las políticas como a los requerimientos de información, es raro, dando como resultado que la evaluación se lleve a cabo, generalmente, de forma apresurada y contraproducente. La comunidad de evaluación británica es limitada en tamaño y experiencia y es empleada principalmente bajo las órdenes del gobierno central. Mientras que hay poca o ninguna evidencia de que exista colusión entre los evaluadores y la propaganda gubernamental, muchos académicos

se han alejado del compromiso activo, ya sea en el análisis de las políticas o en su debate. Como resultado, la calidad de la evaluación y de la formulación de las políticas y el nivel del discurso político ha mermado.

Construyendo una cultura de evaluación de políticas

El intento realizado por el Reino Unido de crear una tendencia en la formulación de las políticas hacia la evaluación ha sido tanto formativo como tranquilizador. Es importante reconocer, sin embargo, que el objetivo de la crítica es el de identificar las posibilidades de mejora que pueden tender a exagerar lo negativo.⁸ Es por lo tanto importante en esta sección de cierre acentuar el mensaje positivo, antes de reflexionar sobre algunos de los pre requisitos para alentar un uso más efectivo de la evidencia de evaluación.

La evaluación de las políticas podría proporcionar información en todas las etapas del proceso de la política desde la pregunta “¿cuál es el problema?” a “¿qué tan bien lo hicimos? Cada vez se hace más frecuente, y afortunadamente ocurre de manera creciente. En tanto la evaluación es acertada, probablemente la formulación de las políticas mejorará y los recursos se utilizarán mejor para el beneficio colectivo. Por otra parte, tomando a Gran Bretaña como un estudio de caso, es evidente que con el compromiso político posiblemente se pueda colocar a la evaluación en el centro de la formulación de las políticas y hacer un uso intenso de los resultados de la investigación de evaluación.

Lo que es menos claro del estudio de caso es, si la nueva cultura británica tanto de políticas de evaluación como de evidencia surgida de la evaluación está asegurada y si es capaz de conseguir mejorías en la calidad y en la cantidad, así como asegurar que se utilice la evidencia más sólida. En términos generales, se debe preguntar si los principios que subyacen a las buenas evaluaciones son compatibles con las realidades del mundo de las políticas y si un grado de convergencia es posible para crear un nuevo paradigma sostenible de políticas.

Regresando a la necesidad de mejorar, la revisión anterior y el estudio de caso señalan hacia cinco vías en las cuales el desarrollo de una cultura de evaluación podría ser promovida. Primero, no existe un único mejor modelo de investigación de evaluación. Mientras que, por ejemplo, la experimentación utilizando la asignación aleatoria puede tener algunos atractivos, será inapropiada en diversas ocasiones. Sin embargo, el calendario en relación con el ciclo de la política es siempre de una importancia central; la evaluación solamente puede ganar en el proceso de la política si los resultados están disponibles cuando se requieren. El desafío es establecer la infraestructura necesaria para explotar diferentes modelos de evaluación que respondan a las preguntas relevantes en cada fase del ciclo de la política.

Segundo, para ser más efectivo, la evidencia de la evaluación tiene que estar en el centro del proceso de la política pero, tan distante como sea posible de la arena política. Si está contaminado por la ideología, o si cuando menos se piensa que lo está, pierde su poder, llevando consigo no más peso que una opinión personal y difiriendo poco de la propaganda.

Tercero, las políticas del mundo real son complejas en sus objetivos e implementación. Se sitúan en un ambiente de la política que, cuando menos, es aún más complicado. Tanto las políticas como el ambiente están caracterizados por cambios rápidos y a menudo impredecibles. Todo esto sugiere que los evaluadores podrían necesitar siempre de hacer intercambios entre la precisión y la solidez, y los formuladores de las políticas aceptar que la evaluación trata sólo con modelos de las políticas, representaciones simplificadas de las políticas que comparten sólo algunas de las propiedades de la realidad.

Cuarto, en los avanzados estados de bienestar, el impacto de las innovaciones en las políticas puede ser probablemente marginal —los efectos más grandes probablemente ya se alcanzaron a través de la introducción de las políticas públicas que definieron el surgimiento de la democracia de representación y del estado de bienestar—. Casi con certeza, los efectos serán menores que aquellos pronosticados por los propugnadores del cambio. Esto puede, en ocasiones, hacer que la evaluación acumulativa sea prohibitivamente cara y que requiera apoyarse en el seguimiento y en los enfoques formativos.

Finalmente, la evaluación de las políticas es un arte extremadamente difícil que requiere de conocimientos sustantivos, entendimiento práctico, experiencia metodológica y sobre todo, creatividad. Existe la necesidad sistemática de reunir experiencia y de atraer nuevos talentos.

Tiempos y estrategia

La clave para desarrollar una cultura de evaluación es la estrategia y la clave de las estrategias son los tiempos. Con los horizontes de las políticas mucho más reducidos que aquellos asociados con la evaluación, es imprescindible establecer canales de información antes de demandar la información.

Esto requiere de la creación y administración de una completa infraestructura de la política de información. Para integrar tal infraestructura se requeriría contorsión:

- Una continua y sistemática ampliación de las necesidades futuras de información;
- Una permanente recolección, recuperación y análisis acumulativo de la información administrativa y de seguimiento de la política actual;

- El desarrollo de datos sociales longitudinales y de modelos que faciliten los intentos para predecir los futuros problemas de las políticas y sus soluciones; y,
- La construcción de bases totales y accesibles de datos sobre estudios de evaluación e investigación que faciliten la revisión sistemática de la evidencia.

Este enfoque apuntaría, tanto como sea posible, a reducir el elemento de sorpresa en la formulación de las políticas a través de la construcción de horizontes más amplios en el tiempo. Lo que colocaría al seguimiento, a la revisión sistemática, al análisis secundario y al meta-análisis en el centro de la estrategia de evaluación, pero se necesitaría combinar con la acumulación del análisis y el desarrollo teórico del problema, más que de una orientación hacia una política. El hacer que los datos administrativos y de la política estén disponibles para la comunidad académica podría estimular proactivamente a ésta última hacia el financiamiento de las políticas, así como hacia la investigación relevante sobre las políticas.

La información y el análisis estarían disponibles para capitalizar el uso sistemático de la experiencia de la política y sus metáforas responderían a las cuestiones emergentes de la política. Esto derivaría de una evaluación prospectiva excelentemente afinada utilizando técnicas tales como la microsimulación, los juegos y la experimentación. La implementación de la política estaría generalmente acompañada por prototipos o estudios de implementación diseñados para describir y entender los modos en los cuales las políticas operaron. Esto se beneficiaría si se diseña en sus propios términos en lugar de, como en la experiencia británica, imita los análisis de impacto. Ocasionalmente se pueden explotar las ventajas de un horizonte más extenso en el tiempo, que surja de la investigación sobre las necesidades de información, las evaluaciones de los programas se podrían utilizar para modelar y probar los principios de las políticas.

Una estrategia exitosa de la política basada en las evidencias también estimularía su demanda, la utilización de la evidencia de evaluación y su producción. Esto requiere priorizar la evidencia de la evaluación en el proceso de la política mediante la regulación, la creación de expectativas, el aseguramiento del financiamiento y el entrenamiento del personal a cargo de la política. Los diseñadores de la política podrían requerir de una guía, tanto al delimitar sus peticiones de evidencia, como en su interpretación y utilización. Esto requeriría de capacitación adicional (se espera que todos los funcionarios civiles superiores en el Tesoro del Reino Unido tengan conocimientos de economía) o de un reclutamiento especializado de personal con los antecedentes adecuados para trabajar con la política o con los diseñadores de la política. Los políticos, especialmente quienes son parte del ejecutivo, podrían también beneficiarse de la enseñanza sobre la producción y la naturaleza de la evidencia y su potencial

contribución al proceso de la política. (En el Reino Unido, los nuevos ministros reciben algún entrenamiento de inducción de este tipo.)

En resumen, el ideal razonable es que la representación esquemática de la Tabla 3.1 sea transformada en realidad, asegurando que las diversas formas de evaluación estén disponibles antes de tiempo para proporcionar información durante los diversos momentos del ciclo de la política.

Distanciando la evaluación de la política

El diseño, ejecución e interpretación inicial de las evaluaciones de la política se llevan a cabo de mejor manera, por expertos que no están expuestos a presiones políticas. Es, por lo tanto, importante que los sistemas administrativos creen murallas chinas entre los políticos y los evaluadores. En algunos países, la legislación y la práctica protegen la recolección y la emisión de las series estadísticas gubernamentales de la interferencia política, con ello aseguran que todas las circunscripciones puedan confiar en los datos que utilizan. El mismo modelo debería aplicarse a las evaluaciones de la política.

Alcanzar esta posición podría ser especialmente difícil en circunstancias donde las políticas son controversiales o donde los políticos tienen un interés personal cercano. El desafío es establecer distancia sin reducir el papel, la eficacia y la influencia de la evaluación. Colocar la responsabilidad exclusivamente en las manos de organizaciones independientes hace difícil el asegurar que los resultados lleguen a los diseñadores de las políticas cuando éstos sean más necesarios con lo que se incrementa la probabilidad que los matices de los problemas de la política se pasen por alto y se reduzca el acceso de los evaluadores a los datos administrativos por razones de confidencialidad. Una estrategia más efectiva podría ser la combinación de procedimientos de rutina para iniciar las evaluaciones y recolectar las conclusiones conjuntamente con un papel claro para el escrutinio externo de los métodos, de la interpretación y de la publicación.

La independencia y la libertad de la interferencia política son importantes por una razón ética adicional que es más evidente en relación con las evaluaciones de los programas y los prototipos. Contrario a las investigaciones ordinarias de las políticas, que son típicamente pasivas, la evaluación de los programas busca explícitamente cambiar el comportamiento de las personas para probar la validez de un modelo de política. Mientras que las políticas implementadas por los cuerpos gubernamentales están, por lo general, dirigidas directamente para la promoción del bien común, la evaluación de los programas y la experimentación logran esta meta sólo indirectamente; el propósito principal es simplemente averiguar qué pasa cuando se introduce una política. Más aún, implícita en la noción de la evaluación del programa está la posibilidad que las cosas vayan mal. La política puede no trabajar o no funcionar tan bien como se esperaba o tan efectivamente como otra

política. La evaluación es, por lo tanto, un ejercicio inherentemente riesgoso que podría potencialmente causar “daño” a quienes participan en el experimento; pueden perder ingreso, por ejemplo, o sufrir consecuencias “psicológicas” tales como tensión y pérdida de la autoestima (Newman y Brown, 1996).

“El argumento ético a favor de la experimentación es que es preferible infligir un daño posible a pequeña escala antes que, sin saber, herir en una escala más grande como resultado de una política mal guiada” (Burtless, 1995). En estas situaciones, los investigadores y los diseñadores de políticas deben sopesar las posibles consecuencias negativas de la evaluación frente a las ganancias que se han anticipado. Estas decisiones necesitan tomarse con referencia al bien común tal vez formalmente por un comité independiente de ética. Ciertamente, ellos no deberían ser cautivos de ninguna conveniencia política.

Las premisas éticas que sustentan la evaluación de las políticas también tienen implicaciones en la utilización de los resultados. Una evaluación debe ser socialmente importante para justificar el poner a las personas en riesgo de daño, lo cual a su vez requiere que se preste una total y apropiada atención a los resultados para que éstos proporcionen información al desarrollo de la política. Considerar los resultados como inútiles mina la legitimidad del modelo de evaluación.

Complejidad y realidad

Las complejidades de las políticas, conjuntamente con la expectativa de los efectos a pequeña escala, tienen importantes implicaciones tanto para los diseñadores de las políticas como para los evaluadores. Limitan la precisión con la cual los resultados de la política pueden ser medidos y podrían, además, significar que la actuación frente a ciertos objetivos, algunas veces importantes, no puede ser diagnosticada cuantitativamente. Esto significa que la evaluación de las políticas jamás removerá todos los riesgos asociados con la formulación de la política y que los evaluadores no tienen jamás la probabilidad de heredar el manto de Platón como los gobernantes de la filosofía. Más bien, los evaluadores continuarán proporcionando evidencia para el proceso de la política (con la esperanza que se requiere más frecuente y adecuadamente) que continuará siendo juzgada y utilizada (aunque por los formuladores de las políticas quienes están mejor informados sobre el proceso de evaluación) al lado de otras fuentes de información.

Las mismas consideraciones de complejidad y dificultad de la medición también refuerzan la tendencia a realizar evaluaciones con métodos múltiples. Puede necesitar además un rediagnóstico de la prioridad concedida a la medida cuantitativa y al entendimiento cualitativo, ya que usualmente se darán los intercambios necesarios entre los dos. (En Gran Bretaña, la prioridad usualmente se le ha otorgado a la cuantificación, algunas veces con consecuencias desalentadoras.)

Tan importante como lo anterior es que diferentes métodos sean efectivamente integrados dentro de un diseño. Esto a menudo no ocurre, en gran parte porque la integración de los métodos cuantitativos y cualitativos es extremadamente difícil de alcanzar (Ritchie y Lewis, 2003). A cambio, se implementan diferentes métodos en paralelo y se reúnen en un informe final antes que ser utilizados para definir el enfoque y el contenido de la evaluación (Walker, 2001). El resultado es que las interpretaciones hechas en referencia a los varios componentes de una investigación retienen el estatus de una hipótesis con contenido más que cualquier otro elemento sustancial. Parte del problema, como se anotó antes, es que los diferentes métodos se toman de diferentes disciplinas; quienes las practican son entrenados separadamente y no están al día con el potencial y limitaciones de otros métodos. En las evaluaciones a gran escala, la experiencia en investigación cuantitativa y cualitativa puede también dividirse entre diferentes organizaciones, lo cual requiere de un fuerte liderazgo administrativo y de sistemas que aseguren un continuo enlace entre todos los involucrados.

Pero, el desafío de la integración se extiende más allá de los aspectos prácticos y a menudo requiere de negociaciones a través de las fronteras existentes entre las posiciones ontológicas. La mayor parte de los investigadores involucrados en la evaluación de las políticas probablemente estaría de acuerdo que existe una realidad a ser descubierta más allá del entendimiento subjetivo. Sin embargo, no todos los investigadores, y probablemente menos los diseñadores de políticas y los políticos, aceptarían que existen múltiples realidades sociales, definidas en términos de los puntos estratégicos desde los cuales la realidad es observada. Hasta ahora, cuando se triangula entre varios métodos a menudo es muy difícil reconciliar los diferentes resultados si el concepto de una sola realidad se adhiere a éstos. Esto es especialmente así cuando los resultados de sólo dos métodos están siendo comparados. Si difieren, usualmente se presume que esto es así porque los resultados producidos por uno o el otro o por ambos métodos son erróneos, pero, generalmente, no existe modo de distinguir cuál. En tales circunstancias, los diseñadores de políticas están comprensiblemente propensos a la desesperanza y los evaluadores están tentados a evitar el problema contando historias que los dejen a salvo. Existe una urgente necesidad de traer más y más a la luz estas dificultades de interpretación y de intercambiar experiencias y conocimiento como las formas en que ellas podrían ser abordadas. Si es posible que dos resultados aparentemente conflictivos resultan estar en lo correcto, existe la necesidad de debatirlos así como las circunstancias y condiciones donde esto puede ocurrir.

Expandiendo fuerzas

Esta última observación apunta a la necesidad de expandir la comunidad de académicos y comentaristas con un interés en la evaluación de las políticas.

Ubicados en la interfase entre el proceso de la política y la erudición, los desafíos enfrentados por los evaluadores están actualmente en la frontera de lo que la ciencia social puede ofrecer a la sociedad.

Aun, desde algunas perspectivas en la academia, los evaluadores se están perturbando con meros rompecabezas más que con problemas sustantivos. Ciertamente, el observador casual de las evaluaciones de las políticas encontrará que se da más atención a la metodología de las ciencias sociales que a la teoría de las ciencias sociales. Sin embargo, éste no es siempre el caso y cuando lo es, algunas veces es explicable en términos de consideraciones prácticas. Por ejemplo, como se anotó arriba, en Gran Bretaña, tanto el proceso de la política como los procedimientos para encargar evaluaciones han discriminado el involucramiento de los académicos.

Sin embargo, si una cultura de evaluación está por nutrirse y sostenerse, es esencial conectar la tecnología de la evaluación con sus raíces teóricas sustantivas y metodológicas. Esto podría suceder si la estrategia para producir la evidencia de evaluación bosquejada arriba se pudiera alcanzar, creando periodos de tiempo más largos para la investigación, haciendo que los datos administrativos estén listos más rápidamente y aceptando que la teoría tiene un rol importante en conceptualizar los problemas de la política y en interpretar las conclusiones.

Finalmente, es también necesario ver más allá de las fronteras ya que la teoría de la evaluación y su método (y evaluadores) probablemente se esparcirán mejor que las políticas evaluadas. La Sociedad Europea de Evaluación ya existe y la conferencia de la OCDE para la cual esta contribución fue originalmente escrita marca una importante iniciativa.

Notas

1. Una situación contrafactual podría ser construida teóricamente si se removiera el derecho de acceder a una política sobre una base aleatoria pero, probablemente, esto haga surgir preocupaciones éticas y políticas. Aun así, el seguimiento de las políticas que busca abordar la primera pregunta y la evaluación retrospectiva diseñada a contestar la segunda, son, en la práctica, más comunes que la evaluación de programas.
2. La experimentación ha sido utilizada en la evaluación para explorar la toma de decisiones por los administradores de primera línea (Huby y Dix, 1992).
3. A menudo, éstas son denominadas preguntas de compensación u opciones.
4. Los problemas de definición hacen imposible un número preciso. Al añadir los proyectos pilotos comisionados en Escocia, probablemente, el total ascendería a 100 o más.
5. Es importante reconocer que el gobierno laborista en Gran Bretaña ha establecido metas para áreas más amplias de la política y que la actuación de los departamentos gubernamentales individuales es diagnosticado en relación con estas metas. El gobierno está comprometido a erradicar la pobreza infantil para 2019 aunque no se ha acordado la definición de pobreza.

6. Esta sección descansa sobre todo en Walker (2001).
7. Al Centro para la Administración y los Estudios de Política en la Oficina del Gabinete se le dio un rol estratégico intragubernamental para abordar algunas de estas cuestiones pero recientemente fue sujeto de una reorganización y de una reducción de tamaño.
8. Ciertamente, la carga de negatividad se da cuando diseñadores de las políticas apuntan en contra de los investigadores de las políticas, cuando, ellos solamente están realizando su trabajo.

Referencias bibliográficas

- ASHWORTH, K. et al. (2002a), *Educational Maintenance Allowance: The first two years, a quantitative analysis*, Londres: Department for Education and Skills, Reporte de Investigación 352.
- , CEBULLA, A., GREENBERG, D. y WALKER, R. (2002b), “Meta-Evaluation: Discovering what works best in welfare provision”, Artículo presentado en la Conferencia Anual de la Sociedad para la Evaluación del Reino Unido, en el South Bank Centre, Londres, 12 Diciembre 2002.
- BERTHOUD, R. (1984), *The Reform of Supplementary Benefit*, Londres: Policy Studies Institute.
- BLONDEL, S., LOHÉAC, Y. y RINAUDO, S. (2000), “Rational Decisions of Drug Users: an experimental approach”, reporte preeliminar, The 20th Arne Ryde Symposium on Experimental Economics.
- BREWER, M., CLARK, T., y MYCH, M. (2001), *Credit Where It's Due: An assessment of the new tax credits*, Londres: Institute for Fiscal Studies.
- BRICE, G.C., GOREY, K.M., HALL, R.M. y ANGELINO, S. (1996), “The STAYWELL Program – Maximising Elders’ Capacity for Independent Living Through Health Promotion and Disease Prevention Activities”, *Investigación en curso*, 18, 2, 202-218.
- BULMER, M. (1986), *Social Science and Policy*. Londres: Allen y Unwin.
- BURTLESS, G. (1995), “The Case for Randomised Field Trials in Economic and Policy Research”, *Journal of Economic Perspectives*, 9, 2, 63-84.
- y ORR, L. (1986), “Are Classical Experiments Needed for Manpower Policy?”, *Journal of Human Resources*, 21, 4, 606-639.
- BOTTOMLEY, D. y WALKER, R. (1996), “Experimental Methods for Policy Evaluation”, Loughborough: Centre for Research in Social Policy, Documento de trabajo 276S.
- CARTER, N. y GREER, P. (1993), “Evaluating agencies: Next steps and performance indicators”, *Public Administration*, 71, (4), pp. 407-16.
- CABINET OFFICE (1999), *Professional Policy Making for the Twenty First Century*, Londres: Oficina del Gabinete: Equipo de diseño de políticas públicas estratégicas
- (2000), *Adding It Up: Improving analysis and modelling in central government*, Londres: Oficina del Gabinete, Unidad de Desarrollo e Innovación.
- CM. (1999), *Modernising Government*, Londres: Command Paper, Cm. 4310.
- (2002), *Opportunity For All*, Cuarto reporte Anual, Londres: Cm. 5598.

- CMND 9518, (1985), *Reform of Social Security: Programme for Change*, Londres: HMSO.
- 9691, (1985), *Reform of Social Security: Programme for Action*, Londres: HMSO.
- CHEN, T. (1990), *Theory Driven Evaluation*, Nueva York: Sage.
- CULLEN, J. y HILLS, D. (1996), *The Role of Randomised Controlled Trials in Assessing Effectiveness of Services: A Critical Review*, Londres: The Evaluation Development Review Unit, The Tavistock Institute.
- CURINGTON, W.P. (1994), "Compensation for Permanent Impairment and the Duration of Work Absence", *The Journal of Human Resources*, 29, 888-910.
- DAVIES, H., NUTLEY, S., y SMITH P. (eds) (2000), *Evidence and Public Policy*, Bristol: Policy Press, Septiembre.
- ECKEL, C. y GROSSMAN, P. (2001), "Differences in the Economic Decisions of Man and Women: Experimental Evidence". En C. Plott y V. Smith (eds.) *Handbook of Experimental Results*, Nueva York: Elsevier.
- GLASS G. (1976), "Primary, Secondary and Meta-Analysis of Research." *Education Research*; 5:3-8.
- GIBSON, C. y DUNCAN, G. (2002), "Lessons Learned: The advantages of mixed methods in program evaluation", Artículo presentado en la Conferencia Annual de la: American Association of Policy Analysis and Management, Dallas, 8 de noviembre.
- GOODYEAR-SMITH F. y FARNELL, A. (2001), "Pain Management in Palliative Care: a modified Delphi consultation study", www.rnzcgp.org.nz/NZFP/Issues/Apr01/or_gys.htm.
- GREENBERG, D y SCHRODER, M. (1997), *Digest of Social Experiments*, Washington: Urban Institute Press (Second edition, third expected 2003).
- , MANDELL, M. y WALKER, R. (1999), "Learning from random assignment experiments in the US and Britain". Artículo Presentado en la Association of Public Policy Analysis and Management, Washington, DC, 6 de noviembre. Disponible en el Centre for Research in Social Policy, Loughborough University, Documento de Trabajo 2264.
- , MANDELL, M., y ONSTOTT, M. (2000), "The dissemination and utilization of welfare-to-work experiments in state policymaking", *Journal of Policy Analysis and Management*, 19, 3, 367-382.
- GUESS, G. y FARNHAM, P. (2000), *Cases in Public Policy Analysis*, Washington: Georgetown University Press.
- HALES, J., LESSOF, C., ROTH, W., GLOYER, M., SHAW, A., MILLAR J., BARNES, M., ELIAS, P., HASLUCK, C., MCKNIGHT, A. y GREEN, A. (2000), *Evaluation of the New Deal for Lone Parents: Early Lessons from the Phase One Prototype – Synthesis Report*, Londres: Department of Social Security (DSS), Research Report 108.
- HALL, P. LAND, H. PARKER, R. y WEBB, A. (1975), *Change, Choice and Conflict in Social Policy*, Londres: Heinemann.
- HAMMERSLEY, M. (1992), *What's Wrong with Ethnography*, Londres: Routledge.
- HARRÉ, R. (1986), *Varieties of Realism*, Oxford: Blackwell.

- HASLUCK, C., MCKNIGHT, A. y ELIAS, P. (2000), *Evaluation of the New Deal for Lone Parents: Early lessons from the Phase One prototype – Cost-benefit and econometric analyses*. Londres: DSS Research Report 110.
- HILLS, D., WARD, C. BLACKBURN, Y. y YOULL, P. (2001), *Evaluation of the New Deal for Disabled People: Innovative schemes pilot*. Londres: DSS Reporte de Investigación 143.
- HOOD, C. (1991), "A Public Management for all Seasons", *Public Administration*, 69, (1), pp. 3-19.
- HUBY, M. y DIX, G. (1992), *Evaluating the Social Fund*, Londres: DSS Reporte de Investigación.
- KLEVMARKEN, N. A. (1997), "Behavioral Modelling in Micro Simulation Models", Uppsala: Department of Economics, Documento de Trabajo, 31. Con acceso a través del vínculo: www.nek.uu.se/Pdf/1997wp31.pdf, 28 de Octubre de 2002.
- KNAPP, M. (1984), *The Economics of Social Care*, Houndmills: Macmillan.
- LAKATOS, I. (1970), "Falsification and the methodology of scientific research programmes" en: I. Lakatos y A. Musgrave (eds.) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.
- LIPSEY, M. y WILSON, D. (2001), *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- LISSENBURGH, S. (2001), *New Deal for the Long-term Unemployed: A comparison of provision in pilot and national areas*. Sheffield: Servicio para el Empleo, Reporte de Investigación y Desarrollo.
- LOUMIDIS, J., STAFFORD, B., YOUNGS, R., GREEN, A., ARTHUR, S., LEGARD, R., LESSOF, C., LEWIS, J., WALKER, R., CORDEN, A., THORNTON, P. y SAINSBURY, R., (2001), *Evaluation of the New Deal for Disabled People Personal Adviser Pilot*, Londres: DSS Reporte de Investigación 144.
- MICHALOPOULOS, S. et al. (2002), *Making Work Pay: Final report on the Self-sufficiency Project for Long-term Welfare Recipients*, Ottawa: Social Research and Demonstration Corporation.
- NEWMAN, D. y BROWN, R. (1996), *Applied Ethics for Program Evaluation*, Thousand Oaks: Sage Publications.
- PATTON, M. (2002), *Qualitative Research and Evaluation Methods*, Thousand Oaks: Sage Publications.
- PAWSON, R. y TILLEY, N. (1997), *Realistic Evaluation*, Londres: Sage.
- PLUM, G. (1998), accessed from www.uni-klorenz.de/~kgt/Learn/Textbook, 28 de octubre de 2002.
- ORR, L. (1998), *Social Experiments Evaluating Public Programs With Experimental Methods: Evaluating Public Programs with Experimental Methods*, Nueva York: Sage.
- ROGERS, P., HACSI, H., PETROSINO, A. y HUEBNER, T. (2000), *Program Theory in Evaluation: Challenges and opportunities, new directions for evaluation*, 87, San Francisco: Jossey-Bass.
- RITCHIE y LEWIS, J. (2003), *Qualitative Research Practice*, Londres: Sage.

- SHAW, I. (2000), *Evaluating Public Programmes: Contexts and issues*, Aldershot: Ashgate.
- SMITH, A., YOUNGS, R., ASHWORTH, K., MCKAY, S. y WALKER, R., (2000), *Understanding the Impact of Jobseeker's Allowance*, Londres: DSS Research Report 111.
- STAFFORD, B. GREENBERG, D y DAVIS, A. (2002), *A Literature Review of the Use of Random Assignment Methodology in Evaluations of US Social Policy Programmes*, Londres: Department of Work and Pensions, In-house Report, 94.
- STEWART, J., (1996), "A Dogma of our Times – The Separation of Policy-Making and Implementation", *Public Money and Management*, Julio - Septiembre, 33-40.
- THOMAS, P. (1987), "The Use of Social Research: Myths and models", pp. 51-60 en M. Bulmer (ed.) *Social Science Research in Government*, Cambridge University Press.
- THORNTON, P. y CORDEN, A. (2002), *Evaluating the Impact of Access to Work: A case study approach*, Sheffield: Reporte de Investigación y Desarrollo WAE138, Claimant Unemployment and Disadvantage Analysis Division, Department for Work and Pensions.
- VOYER, J-P., JOHNSON, K., MONTMARQUETTE, C. y ECKEL, C. (2002), *Fostering Adult Education: A laboratory experiment on the efficient use of loans, grants and savings incentives*, Ottawa: Social Research Demonstration Corporation
- WALKER, D. (2000), "You Find the Evidence, We'll Pick the Policies", *The Guardian*, 15 de Febrero, p3H.
- WALKER, R. (2004), *Social Security and Welfare*, Milton Keynes: Open University Press, próxima publicación.
- (2003), "Applied Qualitative Research" en A. Bryman (ed.) *Encyclopedia of Social Science Methods*, Londres: Sage.
- (2001), "Great Expectations: Can Social Science Evaluate New Labour's Policies?" *Evaluation*, 7, 3, 305-330.
- (2000a), "Learning if Policy Will Work: The case of the New Deal for Disabled People" *Policy Studies*, 21, 4, 313-345.
- (2000b), "Welfare Policy: Tendering for evidence", pp. 141-166 in H. Davies, S. Nutley y P. Smith (eds.) *Evidence and Public Policy*, Bristol: Policy Press.
- (1997), "Public Policy Evaluation in a Centralised State", *Evaluation*, 3, 5, pp. 261-279.
- (1987), "Perhaps Minister: The messy world of 'in-house' Government research", pp.141-165. in Bulmer, M. (ed.), *Social Science Research in Government*, Cambridge University Press.
- (ed.) (1985a), *Applied Qualitative Analysis*, Aldershot: Gower.
- (1985b), *Housing Benefit: the experience of implementation*, Londres: Housing Centre Trust.
- y WILLIAMS, J., (1987), "Housing Benefit: Some determinants of administrative performance", *Policy and Politics*, 14, 3, 309-34.
- , HEDGES, A. y MASSEY, S. (1987), *Housing Benefit: Discussion about reform*, Londres: Housing Centre Trust.

- WEISS, C. (1980), "Knowledge Creep and Decision Creep", *Knowledge: Creation, Diffusion, Utilisation*, 8, 2, 274-81.
- (1992), *Organisations for Policy Analysis: Helping government think*. Newbury Park: Sage Publications.
- (1996), "Theory Based Evaluation: Past, present and future", Paper presented at the American Evaluation Association Conference Atlanta, Georgia 8 de noviembre.
- (1998), "Have we learned anything new about the use of evaluation?", *American Journal of Evaluation*, 19, 1, 21-33.
- WILENSKY, H. (1997), "Social Science and the Public Agenda: Reflections on the relationship of knowledge to policy in the United States and abroad", *Journal of Health Politics, Policy and Law*, 22, 5, 1241-65.
- WHITE, M. y RILEY, R. (2002), *Findings from the Macro evaluation of the New Deal for Young People*, Londres: Department for Work and Pensions.
- WILLIAMS, W. (1998), *Honest Numbers and Democracy: Social policy analysis in the White House, Congress and the Federal agencies*. Washington: Georgetown University Press.
- WORTHEN, B. SANDERS, J., FITZPATRICK, J. (1996), *Program Evaluation: Alternative approaches and practical guidelines*, Boston: Addison-Welsey.
- YANOW D. (1999), *Conducting Interpretive Policy Analysis*, Thousand Oaks: Sage: Qualitative Research Methods Series, 47.

Capítulo 4

La evaluación del impacto de las políticas de desarrollo económico local en los resultados de la economía local: ¿qué se ha hecho y qué se puede hacer?

Por
Timothy J. Bartik
*Economista Senior del W. E. Upjohn Institute for Employment Research,
Kalamazoo (EUA)*

Este artículo sostiene que es factible realizar evaluaciones rigurosas de las políticas de desarrollo económico local. Los programas que apoyan a pequeñas empresas selectas se pueden evaluar rigurosamente utilizando un enfoque experimental asignando de manera aleatoria a algunas empresas para que reciban estímulos intensivos en la promoción del programa. Los programas que ayudan a las localidades marginadas pueden ser evaluados rigurosamente asignando de manera aleatoria el programa entre las zonas marginadas elegibles. Si un experimento no puede llevarse a cabo, pueden utilizarse una variedad de enfoques estadísticos para comparar a las firmas o a las zonas que utilizan el programa con grupos de comparación de firmas o zonas que no utilizan el programa. Estos análisis estadísticos deberían ser complementados con encuestas y grupos focales de empresas que utilizan el programa, lo cual proporciona percepciones de porqué el programa funciona o no. Las evaluaciones deberían ir más allá de los efectos de los programas sobre el crecimiento empresarial y centrarse en los efectos sobre la salud fiscal local y los ingresos de los desempleados. Las evaluaciones que utilizan enfoques rigurosos indican que proporcionar servicios de información a los pequeños fabricantes suele ser eficaz. Los programas dirigidos a zonas marginadas son ineficaces a menos que se utilicen grandes cantidades de recursos durante largos periodos.

Prefacio

Este artículo de investigación propone que las políticas de desarrollo económico local pueden ser evaluadas con mayor rigor y así debería ser. La evaluación debería intentar determinar el impacto de las políticas públicas en los resultados de la economía local; es decir, en qué difieren los resultados económicos locales comparándolos con lo que habría sucedido “sino fuera por” las políticas.

Los programas que proporcionan servicios de ayuda financiera a las pequeñas y medianas empresas (PyMEs) pueden ser evaluados rigurosamente utilizando métodos experimentales. En dichos experimentos aleatorios, el programa sería promovido selectivamente en PyMEs elegidas al azar, el grupo “de tratamiento”, mientras que un grupo de control compuesto por PyMEs sería elegible para los servicios pero no recibiría esfuerzo especial alguno de promoción. El impacto

de las políticas en las PyMEs se puede evaluar comparando los resultados económicos y el uso del programa en los grupos de tratamiento y de control.

Los programas que apuntan al apoyo de zonas empresariales en localidades marginadas también se pueden evaluar rigurosamente utilizando métodos experimentales. Las zonas designadas para recibir la asistencia pueden ser seleccionadas al azar entre las localidades con problemas, y los escasos fondos disponibles pueden ser concentrados con mayor intensidad en estas zonas de “tratamiento” elegidas al azar. El impacto de la política en las zonas designadas se puede evaluar comparando los resultados económicos en estas zonas con los de las zonas marginadas que no fueron elegidas al azar para recibir el apoyo.

Si no hay datos experimentales o el experimento no es factible, los programas de desarrollo económico local pueden y deben ser evaluados mediante análisis estadísticos de los resultados en las empresas o zonas que utilizan los programas o que utilizan los programas más intensivamente (el grupo “de tratamiento”) y los resultados económicos en empresas o zonas de comparación (el grupo “de comparación”). Existe una variedad de técnicas estadísticas bien desarrolladas que intentan determinar las diferencias en los resultados económicos entre los grupos de tratamiento y de control que se pueden atribuir al programa.

Estos análisis estadísticos se deben complementar con encuestas y grupos de enfoque dirigidos a las empresas usufructuarias de los programas de desarrollo económico. Las encuestas que se administran de manera independiente, que garantizan el anonimato e indagan cuestiones específicas, pueden proporcionar evidencia adicional sobre la eficacia del programa para influir en las prácticas de las empresas. Las encuestas y los grupos de enfoque pueden proporcionar también algunas ideas sobre cómo y porqué es eficaz el programa e indicar cómo puede mejorarse.

Las evaluaciones deben tratar de ir más allá del impacto de las políticas en el incremento del crecimiento empresarial local y apuntar a los beneficios de las políticas para el público. Estos beneficios incluyen los ingresos fiscales para el gobierno y mayores ingresos para los desempleados o subempleados. Los beneficios fiscales y del empleo se pueden estimar utilizando modelos econométricos regionales que se combinan con módulos especiales que toman en consideración la estructura de los impuestos y los presupuestos gubernamentales locales, así como el mercado laboral local.

En los Estados Unidos, estos enfoques de evaluación más rigurosos se han aplicado ampliamente, pero de ningún modo universalmente, a organizaciones federales, estatales y locales relacionadas con el desarrollo económico. Los resultados de estas evaluaciones sugieren que los programas de desarrollo económico que proporcionan información, capacitación y servicios de consultoría a pequeños y medianos fabricantes son a menudo efectivos para mejorar el

desempeño local de las empresas. Sin embargo, los programas que están dirigidos a zonas marginadas, como las zonas empresariales, tienden a no ser efectivos si los servicios y la ayuda financiera ofrecidos son demasiado modestos como para compensar las desventajas económicas de las zonas marginadas; los programas de desarrollo económico más efectivos para las zonas marginadas, como la Comisión Regional de los Apalaches, movilizan gran cantidad de recursos por un largo periodo. Los programas de desarrollo económico tienen a menudo significativos beneficios fiscales y para el empleo, el grado de estos beneficios varía ampliamente dependiendo de las condiciones locales. Los modelos pueden estimar estos beneficios fiscales y para el empleo si los modelos incorporan los efectos de las condiciones especiales locales.

Alentar la evaluación más rigurosa de las políticas de desarrollo económico local probablemente requiere la intervención de niveles más altos de la esfera gubernamental. Estos niveles superiores del gobierno deberían proporcionar financiamiento para las evaluaciones y requerir las evaluaciones cuando financien programas locales. Tal intervención de las altas esferas del gobierno es necesaria y apropiada, porque los beneficios de evaluar un programa en particular van más allá de la organización que opera el programa y atañen a todas las organizaciones que operan o considerarían operar programas similares y al público.

Introducción

Este artículo de investigación considera los mejores enfoques para evaluar los impactos que tienen las políticas sobre los resultados económicos locales deseados.¹ El artículo se apoya principalmente en el conocimiento que poseo de las políticas para el desarrollo económico estatal y local en los Estados Unidos, se presentan situaciones similares al evaluar políticas de desarrollo económico locales en otros países de la OCDE.

Este artículo de investigación trata de responder a nueve preguntas:

- ¿Cuáles son los programas de desarrollo económico que estamos tratando de evaluar y por qué son importantes?
- ¿Qué tipo de evaluación de estos programas es el más necesario?
- ¿Qué sesgos aparecen al evaluar estos programas?
- ¿Podemos utilizar eficazmente experimentos aleatorios para evaluar los programas de desarrollo económico?
- ¿Podemos utilizar métodos estadísticos para hacer grupos de comparación no aleatorios realmente comparables?
- Si una zona local tiene un enfoque de desarrollo económico que es realmente “único”, ¿éste puede ser evaluado?

- ¿Además de la comparación estadística con grupos de control o de comparación, hay alguna otra evidencia que pudiera indicar el impacto del programa?
- ¿Podemos determinar por qué y cómo un programa tiene impactos o no logra tenerlos?
- ¿Podemos determinar los impactos del programa sobre los objetivos económicos ulteriores y no sólo sobre los inmediatos?

¿Cuáles son los programas de desarrollo económico que estamos tratando de evaluar y por qué son importantes?

Con la denominación “programas de desarrollo económico local” me refiero a programas que proporcionan asistencia a empresas más o menos dirigida a sus necesidades, con la meta inmediata de incrementar la actividad empresarial en la economía local. (Hay, por supuesto, objetivos económicos ulteriores a alcanzar mediante el incremento de la actividad empresarial local que abordaré más adelante.)

Hay muchas maneras de clasificar dichos programas de desarrollo económico local. La Tabla 4.1 contiene un sistema de clasificación que organiza los programas de un modo que es pertinente para las técnicas de evaluación. Los programas de desarrollo económico local del primer tipo son aquellos que prestan servicios o asistencia financiera a sólo algunos candidatos elegibles, usualmente pequeñas y medianas empresas (PyMEs), y los candidatos se autopropone para recibir asistencia o son seleccionadas por los programas. Este tipo de servicios o de asistencia financiera puede incluir información o capacitación para los gerentes o trabajadores de la empresa, o apoyo financiero público para la puesta en marcha o la expansión de la empresa.

Un segundo tipo de programa proporciona apoyo financiero o servicios a todas las empresas ubicadas en una zona específica, la cual ha sido designada como marginada por algún nivel superior del gobierno que ayuda a financiar el programa. Los ejemplos en los Estados Unidos abarcan programas de zonas empresariales patrocinados por muchos gobiernos estatales, como el programa de la “Zona de Potenciación” instaurado por el gobierno federal durante la administración Clinton, y la Comisión Regional de los Apalaches iniciada en los sesentas.

Un tercer tipo de programa proporciona apoyos en toda la zona que patrocina el programa y a todas, o casi todas, las empresas elegibles para el apoyo, aunque a menudo los lineamientos de elegibilidad de las empresas dirigen los apoyos hacia el tipo de empresas que se considera que proporcionarán los mayores beneficios económicos. Un objetivo común de tales programas son las empresas de manufactura u otras firmas “orientadas a la exportación” que venden sus productos fuera de la zona que patrocina el programa, aunque en ocasiones

los programas se enfocan hacia una industria en particular, tales como algunas industrias de alta tecnología. Estos programas incluyen: promover una zona como ubicación para nuevas instalaciones corporativas; ayudar a resolver problemas con la regulación gubernamental para nuevas instalaciones o expansión de las mismas; proporcionar beneficios fiscales, infraestructura en una ubicación específica o adecuar a las necesidades la capacitación de nuevos trabajadores para abrir o expandir instalaciones; y trabajar con las redes o *clusters* de negocios en una zona para mejorar los servicios o la infraestructura local.

Tabla 4.1 **Clasificación de las políticas de desarrollo económico local**

<p>1. Apoyos a firmas seleccionadas (predominantemente a pequeñas y medianas empresas)</p> <ul style="list-style-type: none"> ● Capacitación para emprender o administrar una empresa ● Préstamos/inversión pública o apoyo público para préstamos/inversión privada para el inicio de empresas o su expansión ● Información/capacitación sobre la implementación de nueva tecnología o nuevas técnicas gerenciales ● Capacitación de nuevos trabajadores adecuada a las necesidades de la firma o industria ● Información/capacitación para la exportación
<p>2. Apoyos a zonas marginadas (zonas empresariales y otros programas diseñados y designados generalmente por los niveles superiores del gobierno)</p> <ul style="list-style-type: none"> ● Beneficios fiscales sobre impuestos municipales y de jurisdicciones gubernamentales para las empresas que se asientan o expanden en la zona designada ● Servicios o infraestructura mejorados en la zona designada, ya sea para empresas específicas o en lo general
<p>3. Programas generales para la zona (dirigidos típicamente a fabricantes u otras empresas "orientadas a la exportación"; en ocasiones dirigidos a ciertas industrias)</p> <ul style="list-style-type: none"> ● Promover una zona y proporcionar información sobre el sitio a prospectos de nuevas plantas de mercado ● Proporcionar ayuda a las empresas existentes y a las nuevas para resolver los problemas de la regulación gubernamental ● Aprovechamiento expedito de caminos y servicios públicos para nuevas plantas o expansiones en sitios específicos o desarrollo previo de parques industriales ● Incentivos fiscales para plantas de mercado nuevas o expandidas, o para sedes corporativas ● Capacitación de nuevos trabajadores adecuada a las necesidades de la empresa como incentivo para crear nuevas instalaciones corporativas o expansiones ● Apoyo a la creación de redes o <i>clusters</i> de negocios en una industria para desarrollar mejores servicios de soporte, como la capacitación ● Giros tecnológicos o industriales a cualquiera de los programas antes mencionados, por ejemplo parques industriales o incentivos fiscales o capacitación orientada a la tecnología

En los Estados Unidos se estima que aproximadamente entre 20 y 30 mil millones de dólares del gasto gubernamental, estatal y municipal, o de los impuestos se dedica anualmente a programas de desarrollo económico "personalizados", con quizás otros 6 mil millones de dólares anuales para apoyos por parte del gobierno federal.² El impresionante monto de dichos recursos se dirige a programas generales de las zonas, en su mayoría en forma de incentivos fiscales. Por ejemplo, un reciente estudio del estado de Michigan indica que, de los 700 millones de

dólares en recursos (unos \$70 per cápita) dedicados anualmente a los programas de desarrollo económico, más de \$600 millones se dedican a programas que operan a lo ancho del estado para casi todas las firmas elegibles, y más de tres cuartas partes de esos \$600 millones se traducen en beneficios fiscales, básicamente reducciones en los impuestos sobre la propiedad para instalaciones de manufactura nuevas o su expansión (Bartik, Eisinger y Erickcek 2003).

Sin embargo, aunque la cantidad de 20 o 40 mil millones de dólares en recursos es bastante significativa, en los Estados Unidos la importancia del desarrollo económico local va más allá de esta relativamente estrecha definición de política pública para el desarrollo económico local. Este tipo de definición restringida se centra en políticas que son claramente hechas a la medida de las necesidades de firmas en particular o que están dirigidas a grupos particulares de firmas, y excluye muchas de las políticas estatales y municipales más generales.³ En los debates estatales y municipales sobre impuestos, gasto o políticas de regulación, los efectos de las políticas públicas sobre el desarrollo económico de las zonas estatales o municipales siempre son una consideración importante (Peterson 1995, 1981). Por ejemplo, en años recientes, casi tres cuartos del total de los estados en los Estados Unidos han cambiado el enfoque de repartirse los ingresos de una corporación multi-estatal entre los estados y han pasado a aplicar un enfoque que basa la mitad o más de la fórmula en la participación del estado en las “ventas” de la corporación, lo cual suele reducir en gran medida la recaudación de impuestos sobre los ingresos corporativos de las empresas que exportan una gran parte de su producto fuera de los límites estatales (Mazerov 2001; McLure y Herllerstein 2002). Este dramático cambio en la política estatal de los impuestos empresariales se racionaliza usualmente como una manera de impulsar el desarrollo económico del estado. El impulso al desarrollo económico se utiliza también para racionalizar muchos otros cambios en las políticas estatales y locales: otros métodos para disminuir los impuestos empresariales del estado o el municipio; menores impuestos a las personas físicas, en particular los que pagan individuos con altos ingresos; reducción del nivel de las prestaciones sociales; cambios en las leyes de compensación a los trabajadores o a los desempleados; y cambios en las regulaciones ambientales o de salud o de seguridad.

Por lo tanto, es importante evaluar las políticas de desarrollo económico local, no sólo por los miles de millones de dólares de recursos involucrados, sino también porque la actividad para el desarrollo económico es claramente una de las más importantes funciones de los gobiernos estatales y municipales en un sistema federal. Distinguir entre aseveraciones duras y débiles sobre los efectos de algunas propuestas de políticas públicas que proporcionan beneficios de desarrollo económico es claramente crucial para tener gobiernos estatales y municipales que funcionen bien.

¿Qué tipo de evaluación de estos programas es la más necesaria?

El tipo de evaluación de las políticas de desarrollo económico local más necesaria es la que consiste en estimaciones del impacto de las políticas sobre los resultados económicos deseables. Denominaré a este tipo de estimaciones evaluación del “impacto resultante”. Idealmente, dicha evaluación debería incluir estimaciones de cómo los impactos resultantes variarían con cualquier cambio posible en el ámbito, escala, diseño o administración de estas políticas o, en otras palabras, que a partir de la evaluación entendamos plenamente cómo y porqué la política pública tiene los impactos estimados. Además, una evaluación ideal no sólo nos diría el impacto de la política en la actividad empresarial local, que es la meta inmediata de las políticas de desarrollo local económico, sino también el impacto de las políticas en el bienestar económico de los residentes locales, meta última de las políticas de desarrollo económico local.

¿Por qué es necesaria la evaluación del “impacto resultante”? Sólo la evaluación del impacto resultante nos proporciona la información requerida si los diseñadores de políticas de hacer una elección informada sobre la opción de la política que maximizará los beneficios sociales.

En los Estados Unidos, una gran cantidad de reportes o estudios pretenden proporcionar “evaluaciones” o “valoraciones del desempeño” de las políticas de desarrollo económico, pero no hacen nada a este respecto. Se ha vuelto cada vez más común que las agencias de desarrollo económico de los estados y municipios produzcan datos considerables sobre las actividades del programa, tales como el número de empleos creados por las firmas asistidas. Los reportes de las agencias a veces afirman que esta creación de empleos es un “impacto del programa”, lo cual equivale a suponer erróneamente que ninguna actividad económica habría ocurrido “de no ser por” el programa de apoyo. Las agencias de desarrollo económico de los estados y municipios reportan también datos sobre las condiciones económicas locales, tales como la creación de empleos durante un periodo en particular o las reducciones en el índice de desempleo. A veces estos reportes establecen las mejoras en las condiciones económicas locales como “impactos del programa”, lo cual equivale a suponer erróneamente que cualquier mejora en la economía local se debe a las políticas de desarrollo económico local.

Por ejemplo, un estudio de “incubadoras de empresas” en los Estados Unidos, que proporcionan espacios a bajo costo, servicios de apoyo compartidos y alguna ayuda de consultoría para iniciar empresas, aseveraba que “los programas de incubación empresarial estudiados en este proyecto han estimulado la creación de miles de nuevos empleos en todo el país” (Molnar *et al.*, 1997, p. 12). El estudio admite que “algunos de los empleos atribuidos a la incubadora se habrían creado incluso sin la existencia de la incubadora puesto que un determinado número de

empresarios siempre emprenderán negocios” (*ibid.*, p. 13). Sin embargo, el estudio asevera que “es imposible saber, después de sucedido el hecho, qué hubiera hecho una empresa sin el apoyo de su programa de incubación empresarial. En este caso, en consecuencia, así como en la mayor parte de la investigación del impacto de los programas de apoyo empresarial, el análisis se centra en el impacto bruto en oposición al neto” (*ibid.*, p. 13). En contraste con las aseveraciones de este artículo de investigación sobre incubación empresarial, yo sostengo que podemos estimar el impacto neto del programa mediante la estimación de lo que habría sucedido, en promedio, si el programa no existiera. Además, yo creo que la terminología “impacto bruto” es engañosa, porque dichas cifras no son necesariamente impactos del programa.

Para evitar la confusión, he de enfatizar que los datos sobre las actividades del programa y las condiciones económicas locales son a menudo útiles. Los datos sobre la actividad del programa ayudan a la gerencia de los programas, y los datos sobre las condiciones económicas ayudan a la comprensión de la economía local. Estos datos pueden ser incluso parte de la información requerida para hacer una verdadera “evaluación de resultados” de las políticas de desarrollo económico local, la cual busca identificar una relación causa-efecto entre las actividades del programa y las condiciones económicas locales, y estimar cuantitativamente su magnitud. Sin embargo, por sí mismos los datos sobre las actividades del programa o las condiciones económicas locales no nos dicen nada de los impactos de las políticas sobre los resultados.

La evaluación del impacto resultante es a menudo costosa por sus exigencias de más datos y pericia en estadística y modelado económico. Debido a que tal evaluación del impacto es costosa, no queda claro que sea necesario ejecutar las evaluaciones en todos y cada uno de los programas operados por cada agencia de desarrollo económico local. Es probable que el mejor consejo para las agencias individuales de desarrollo económico local sea que reserven las evaluaciones del impacto resultante para sus programas más caros, para los cuales son mayores las ganancias posibles de políticas mejor elegidas. Los niveles superiores del gobierno pueden proporcionar un servicio útil pagando la evaluación de programas más pequeños y garantizando que los resultados sean ampliamente difundidos a las agencias de desarrollo económico local que utilizan, o podrían utilizar, programas similares.

¿Qué sesgos surgen cuando se evalúan estos programas?

El estudio ideal —pero imposible— de un programa gubernamental tendría que disponer de la máquina del tiempo de H. G. Wells o de otro escritor de ciencia ficción, retroceder en el tiempo y eliminar el programa sin hacer cualquier otra intervención directa, y después comparar los resultados en este mundo alternativo

inducido sin el programa con los resultados en el mundo original con el programa. A falta de una máquina del tiempo, la mejor alternativa es encontrar algún grupo de entidades que sean comparables con el grupo de entidades que reciben los efectos del programa, pero sin que este grupo de comparación esté involucrado en el programa. Para las políticas de desarrollo económico local del tipo 1 (véase la Tabla 4.1), en el que sólo un subconjunto de empresas elegibles recibe el apoyo, el grupo de comparación consistiría en empresas que no reciben el apoyo. Para las políticas de desarrollo económico local del tipo 2, que se dirigen a zonas marginadas, el grupo de comparación consistiría en zonas que no han sido designadas oficialmente como marginadas. Para políticas de desarrollo económico local del tipo 3, que sirven a todas las empresas elegibles en la zona que patrocina el programa, el grupo de comparación pertenecería a de otras zonas.

Para que dichas comparaciones revelaran inmediata y fácilmente (sin torturas estadísticas) los efectos causales de las políticas de desarrollo económico local sobre los resultados económicos locales, el grupo de comparación tendría que ser el mismo, en promedio, en cuanto a las características observables y no observables que afectan los resultados económicos locales. A falta de datos experimentales, de los que trataremos más adelante, el grupo que recibe el apoyo del programa diferirá en general del grupo de comparación de maneras que afectan los resultados económicos locales. Por lo tanto, es de esperar que el grupo apoyado y el grupo de comparación experimenten variaciones en los resultados económicos, incluso si ninguno de los grupos recibiera el apoyo del programa; en consecuencia, una simple comparación de ambos grupos proporcionará una medición sesgada de los efectos del programa.

¿Cuál es la dirección probable de estos sesgos? Para las políticas de desarrollo económico local que ayudan selectivamente a las empresas (políticas del tipo 1), nuestra intuición es que las empresas de rápido crecimiento son más aptas para la auto-selección a participar en el programa, precisamente porque su crecimiento las lleva a buscar más apoyo financiero y servicios. Hay evidencia sobre la mayor probabilidad de que las empresas de crecimiento rápido utilicen servicios empresariales selectivos proporcionados por las agencias de desarrollo económico local (Jarmin, 1999). Además, existen evidencias de que el crecimiento de las empresas está correlacionado positivamente en el tiempo (Nexus, 1999). En estas condiciones es probable que las empresas que participan en el programa hubieran crecido más rápido en el futuro aun cuando no hubieran participado en el programa, lo cual sesgará las evaluaciones hacia una sobreestimación de los efectos positivos del programa. (Por supuesto que las evaluaciones de los programas de desarrollo económico local en particular pueden tener diferentes sesgos si los programas seleccionan empresas para recibir el apoyo de diferente manera o si el cambio en la variable de los resultados económicos que se examina no está relacionado positivamente en el tiempo.)

Para los programas de desarrollo económico local dirigidos a zonas marginadas (políticas del tipo 2), es probable que estas zonas marginadas —por definición— tengan mayores niveles de miseria económica que las zonas no designadas así (para evidencias, véase Bondonio y Engberg 2000, Greenbaum 1998, Greenbaum y Engberg 1998). Por lo tanto, es probable que esté sesgado hacia el hallazgo de efectos negativos del programa cualquier estudio que compare los niveles de resultados económicos de las zonas objetivo *versus* algún grupo de zonas de comparación, ya que obviamente los niveles de resultados económicos están correlacionados positivamente en el tiempo y por lo tanto las zonas objetivo tendrían mayores niveles de marginación que el grupo de comparación en el futuro sin la intervención del programa. No es tan obvio que las variaciones en los resultados económicos difieran entre las zonas objetivo y las zonas de comparación no designadas. De hecho, algunas evidencias sugieren que en los Estados Unidos, la correlación entre la designación del área como zona empresarial y el crecimiento previo de la zona es insignificante (Bondonio y Engberg, 2000). (Una vez más, las tendencias del sesgo en las evaluaciones de un programa en particular dependerán de las reglas del programa para declarar objetivos.)

Para los programas de desarrollo económico que están al servicio de las firmas elegibles en la zona que patrocina el programa (políticas del tipo 3), las tendencias del sesgo en las evaluaciones dependerán de qué tipos de zonas tengan mayor probabilidad de perseguir agresivamente el desarrollo económico. Las evidencias disponibles indican que en los Estados Unidos, los incentivos tienden a ser más altos en los estados o las ciudades con mayor desempleo y previo crecimiento lento (Fisher y Peters, 1998). Sin embargo, estos incentivos no hacen más que compensar los generalmente, mayores impuestos empresariales básicos efectivos en los estados y municipios que prevalecen en estas zonas de alto desempleo y bajo crecimiento, de modo que la tasa de impuestos empresariales estatales y municipales efectiva después de los incentivos no está fuertemente correlacionada con los índices de desempleo estatales o municipales ni con el crecimiento del empleo. Por lo tanto, el estudio de los efectos de los incentivos podría estar sesgado hacia el hallazgo de efectos sobre el crecimiento económico local de los incentivos menos positivos, ya que sería más probable que las zonas estatales y municipales que hacen uso intensivo de los incentivos crezcan lentamente incluso sin los incentivos. Por otro lado, los estudios que observan los efectos de las tasas básicas de impuestos empresariales estatales y municipales sobre el crecimiento pueden estar sesgados hacia el hallazgo de mayores efectos positivos dados los menores impuestos empresariales, ya que los estados de crecimiento más lento tienden a tener tasas de impuestos empresariales estatales/municipales superiores (para evidencia que confirma en el mismo estado el transcurso del ciclo del negocio, véase Reed y Rogers, 2000).

¿Podemos utilizar con eficacia experimentos aleatorios para evaluar programas de desarrollo económico?

La mejor manera factible de evitar los sesgos al estimar los impactos resultantes de los programas de desarrollo económico es experimentar con los programas mediante la creación de algún proceso aleatorio, el cual ayudará a determinar cuáles entidades (empresas o zonas) utilizarán el programa y cuáles no. Dado que el proceso que determina quien podrá hacer uso del programa es aleatorio, sabemos que los grupos del programa y de tratamiento deben ser los mismos, en promedio, en las variables observables y no observables que afectan los resultados económicos. Cualquier diferencia remanente en los resultados económicos entre los grupos del programa y del tratamiento se deben ya sea al programa, o bien a factores aleatorios que afectan los resultados económicos para una empresa o zona particular. Con un tamaño de muestra suficiente, estos factores aleatorios promediarán cero y estaremos en posibilidad de estimar con precisión el verdadero impacto del programa en los resultados económicos.

Que yo sepa, la única evaluación de desarrollo económico que se ha apoyado en datos generados a partir de un experimento utilizando muestreo aleatorio es un estudio, patrocinado por el Departamento del Trabajo de los Estados Unidos, sobre los efectos de la capacitación empresarial para receptores del Seguro de Desempleo (UI, por sus siglas en inglés) [Benus, Word y Grover 1994]. En este experimento, a los receptores del UI en los estados de Massachusetts y Washington primero se les invitó a sesiones de orientación para explicar el programa de capacitación empresarial. El tres por ciento de los receptores del UI que expresaron interés en este tipo de capacitación después de la orientación fueron asignados al azar a un grupo de tratamiento que recibió dicha capacitación y a un grupo de control que no lo recibió. El 49 por ciento del grupo de tratamiento concluyó con alguna experiencia de autoempleo, en comparación con el 28 por ciento del grupo de control, sin manifestación de que el índice de fracaso empresarial difiriera en los dos grupos. Puesto que el grupo de tratamiento y el de control, en promedio, deberían ser iguales en sus características observables y no observables, podemos confiar en que, salvo por el ruido aleatorio, el 21 por ciento de diferencia en el autoempleo experimentado se debe al programa de capacitación empresarial. Obsérvese que la práctica usual de los programas para solicitar créditos para la actividad empresarial asociada con el programa exageraría los efectos del programa a más del doble, solicitar el crédito para el 49 por ciento del grupo de tratamiento que experimentó el autoempleo. Los programas de desarrollo económico no pueden solicitar legítimamente el crédito por todos los trabajos y demás actividades empresariales que reciben asistencia a través del programa, porque al menos alguna —y quizás todas— la actividades empresariales probablemente habrían ocurrido incluso sin el programa.

Los métodos de experimentación aleatoria podrían aplicarse sin dificultad a otras políticas de desarrollo económico local que sólo apoyan a un grupo selecto de firmas elegibles. Una preocupación relacionada con dicha experimentación es la renuencia a excluir a algunas firmas de los servicios, lo cual sucede en los experimentos clásicos con el grupo de control. Dicha exclusión puede evitarse, si el experimento toma la forma de muestra aleatoria de firmas seleccionadas para una promoción del programa. Los métodos aleatorios se utilizarían para elegir cuáles firmas recibirían un esfuerzo de promoción intensivo, como cartas, llamadas telefónicas y visitas personales, informando a la empresa de los servicios o del apoyo financiero proporcionados por el programa de desarrollo económico. Si esta promoción es suficientemente intensiva, el resultado debería señalar una diferencia significativa en la utilización del programa entre las firmas en el grupo de tratamiento (el grupo objetivo que recibe los esfuerzos de promoción) y las firmas en el grupo de control (el grupo que no recibe los esfuerzos de mercadotecnia). Sin embargo, a ninguna firma en el grupo de control que solicite los servicios se le denegarían arbitrariamente. La diferencia en los resultados económicos (crecimiento en el empleo, mayor productividad, etcétera) entre los grupos de tratamiento y de control de las empresas, dividido por la diferencia en la utilización del programa entre los dos grupos, proporciona una estimación de los efectos del programa. Por ejemplo, consideremos un programa de extensión de la manufactura diseñado para mejorar el crecimiento de la productividad de las empresas y un experimento aleatorio que promueve intensivamente el programa a un grupo de tratamiento de empresas seleccionadas al azar. Si en el grupo de tratamiento la productividad creció 10 por ciento, en el grupo de control la productividad creció cinco por ciento, y la utilización del programa en el grupo de tratamiento fue del 35 por ciento *versus* 10 por ciento en el grupo de control, entonces el efecto estimado de productividad por el programa es de un 20 por ciento de mejora en la productividad $[20 = (10-5) / (0.35-0.10)]$.⁴ Puesto que los grupos de tratamiento y de control difieren en promedio solamente en el número aleatorio que les fue asignado y por lo tanto si el programa fue promovido para ellos, podemos confiar que con un tamaño de muestra suficiente estos cálculos revelarán los impactos del programa sobre los resultados económicos.

Los experimentos aleatorios también pueden ser realizados con los programas de desarrollo económico cuyo objetivo son las zonas marginadas. En general, hay más economías locales afligidas económicamente de las que los niveles superiores del gobierno pueden ser capaces de atender con suficientes recursos para ayudar a dar un giro a la suerte económica de estas zonas marginadas. Además, no está claro si entre las zonas marginadas uno debería enfocarse a las más o las menos afligidas: las zonas más marginadas pueden requerir más apoyo, pero las menos marginadas pueden ser las más fáciles de afectar con el programa adecuado. Por lo tanto, cualquier esfuerzo por parte de los gerentes del programa para seleccionar las zonas objetivo entre todas las zonas marginadas probablemente

reflejaría juicios bastante arbitrarios. Finalmente, en la práctica suele suceder que las jerarquías superiores del gobierno utilizan criterios políticos para seleccionar cuáles zonas marginadas serán designadas para el apoyo. Por ejemplo, durante la administración Clinton, al seleccionar qué zonas de los Estados Unidos serían designadas como “Zona de Potenciación” o “Comunidad Empresarial”, las zonas objetivo finales las eligieron cargos políticos y no se siguió rígidamente el orden desarrollado por un panel de selección. Dada la arbitrariedad inherente y la naturaleza política de los procedimientos actuales para la designación de las zonas marginadas que reciben apoyo, no deberían existir escollos éticos serios para que tal designación se haga utilizando un muestreo aleatorio. Si se hiciera así, las zonas designadas y las no designadas serían, en promedio, iguales en sus características observables y no observables, así como en sus prospectos de crecimiento, y la diferencia en el desempeño económico entre ambos grupos sería una estimación insesgada de los efectos del programa. Para que dichas estimaciones sean lo suficientemente precisas como para ser útiles, tendría que haber un número suficientemente grande de zonas designadas y no designadas de modo que los factores aleatorios sean eliminados en el promedio. Qué tan grande debe ser el tamaño de la muestra depende de que tan grande es el efecto del programa que uno está intentando detectar y de cuánta variación natural hay en el desempeño económico de las zonas marginadas; las metodologías estadísticas estándar permiten contestar sistemáticamente dichas cuestiones del tamaño de la muestra. Como regla general, parece improbable que en la mayoría de los casos pudiera aprenderse mucho sin tamaños de muestra de al menos 20 entidades en cada uno de los grupos de las zonas marginadas designadas y no designadas.

Para las políticas de desarrollo económico local que atienden a de todas las firmas elegibles en toda la zona, la experimentación aleatoria no es posible por definición, ya que estos programas están patrocinados por la zona y el gobierno de la zona no controlará qué programas adoptan los gobiernos de otras zonas. Como ya se mencionó, en los Estados Unidos los programas de áreas reciben la mayoría de los recursos dedicados a las políticas de desarrollo económico local, de modo que para muchas políticas de desarrollo económico la experimentación aleatoria no es factible.

¿Podemos utilizar métodos estadísticos para que grupos no aleatorios de comparación sean verdaderamente comparables?

Debido a que la experimentación no se aplica con frecuencia a las políticas de desarrollo económico local y no es factible con algunas políticas, es importante explorar alternativas. A menudo contaremos con algunos datos sobre los resultados económicos de firmas o zonas que utilizan un programa de desarrollo económico local o que utilizan un programa de manera más intensiva (el grupo de “tratamiento”) y de las que no (el grupo de “comparación”). Debido a que los

grupos de tratamiento y comparación difieren en las variables observadas y no observadas que afectan los resultados económicos, una simple comparación de los resultados entre estos dos grupos puede no revelar los verdaderos efectos del programa. Sin embargo, existe una variedad de técnicas estadísticas que se pueden utilizar para limitar o incluso eliminar los sesgos resultantes de estas diferencias entre los grupos de tratamiento y de comparación.⁵

Este no es el lugar adecuado para profundizar en los detalles técnicos de las técnicas estadísticas apropiadas, pero en suma, existen al menos cinco técnicas estadísticas, no necesariamente excluyentes entre sí, que se pueden utilizar para detectar los verdaderos efectos del programa sobre algunas variables de resultados cuando los usuarios del programa difieren de otras maneras de los no usuarios. Primero, simplemente podemos controlar las variables observadas que afectan el resultado económico, y que podrían estar correlacionadas con el uso del programa, mediante la inclusión de estas variables observadas en la ecuación de estimación que se utiliza para predecir la variable del resultado. Este enfoque es muy eficaz para la reducción del sesgo en la estimación de los impactos del programa cuando tenemos datos sobre todas las variables posibles que afectan el resultado económico de interés y que están correlacionadas con el uso del programa. Este enfoque no puede corregir sesgos que podrían haber sido causados por variables no observables que están correlacionadas tanto con los resultados económicos como con el uso del programa. Este enfoque también parte del supuesto de que conocemos la forma funcional por la cual las variables observables afectan los resultados económicos.

Un segundo enfoque que es una variante del primero se denomina la estimación de la “diferencia en las diferencias”, o estimación de la “diferencia en las diferencias de las diferencias” (DD y DDD para abreviar) (Meyer 1995). Con el enfoque DD comparamos la diferencia antes y después del programa entre los usuarios y los no usuarios del programa o de la política. Con el enfoque DDD, si tenemos alguna razón para pensar que es probable que algunos tipos de usuarios sean más afectados por la política que otros, podemos comparar la diferencia entre los probables grupos de alto impacto y de bajo impacto entre los grupos de usuarios y no usuarios, antes y después de la política. El enfoque DD es equivalente a suponer que se puede hacer una buena labor para controlar otros factores que afectan los resultados económicos permitiendo los efectos del periodo de tiempo, y si la entidad es un grupo de usuarios o de no usuarios, o un grupo de alto impacto o de bajo impacto. La limitación de este enfoque es que pueden existir muchas otras variables, tanto observadas como no observadas, que también afectan los resultados económicos y están correlacionadas con el uso del programa. El segundo enfoque se puede combinar con el primer enfoque agregando algunas de estas otras variables observadas a la ecuación de estimación.

Un tercer enfoque consiste en emparejar usuarios del programa con no usuarios que son similares en las características observadas. La investigación reciente ha revelado que este emparejamiento debería enfocarse a encontrar usuarios y no usuarios que sean lo más similares posible en su “propensión a participar” estimada, que es la probabilidad estimada dadas las variables observadas de que una entidad determinada utilizará el programa (Smith y Todd 2001; Heckman, Ichimura y Todd 1999).

Esta propensión a participar se debe estimar usando variables que predigan la utilización del programa y que tengan una correlación, independiente de la utilización del programa, con los resultados económicos. Las variables que predigan la utilización del programa pero que no predigan los resultados económicos independientemente no se deben incluir en la predicción de la utilización del programa. Este tipo de variables, tanto las observadas como las no observadas, proporcionan la variación en la utilización del programa que es independiente de los factores externos al programa que afectan los resultados económicos.

El enfoque de propensión a participar funciona bien si tenemos datos sobre todas las variables que efectúan una buena labor de predicción sobre la utilización del programa y que también están correlacionadas con los factores externos al programa que afectan los resultados económicos. Este enfoque de emparejamiento no funcionará bien si existen muchas variables no observadas que predicen la utilización del programa y están relacionadas con los resultados económicos. Además, en muchos de los casos puede ser que no existan parejas razonablemente cercanas de algunos usuarios con no usuarios, y por lo tanto las estimaciones de un enfoque de emparejamiento son válidas únicamente como efectos promedio del programa para el tipo de usuarios del programa de los que podemos encontrar buenas parejas entre los no usuarios.

Un cuarto enfoque es modelar explícitamente la selección e integrarla al programa tal como está correlacionada con variables no observadas que afectan los resultados económicos (Murnane, Newstead y Olsen 1985). Para ello, se requiere la estimación de tres ecuaciones: una ecuación que explique los resultados económicos para los usuarios del programa, una segunda ecuación que explique los resultados económicos para los no usuarios, y una tercera ecuación que explique si una entidad dada es usuaria o no del programa. La estimación de la tercera ecuación permite agregar un término de “corrección del sesgo de selección” a cada una de las dos primeras ecuaciones, el cual —en teoría— corrige el sesgo ocasionado por variables no observadas que afectan los resultados económicos y están correlacionadas con la utilización del programa. Este enfoque supone que hemos especificado con exactitud las variables y la forma funcional como deberían ingresar las tres ecuaciones. Adicionalmente, este enfoque supone una distribución estadística particular de los factores no observables (los “términos de error”) que entran en las tres ecuaciones.

Un quinto enfoque requiere encontrar alguna “variable instrumental” que prediga la utilización del programa y no esté correlacionada con las variables no observables que influyen en los resultados económicos (Angrist y Krueger 2001). Con este enfoque de variable instrumental, examinamos solamente las variaciones en los resultados económicos que se pueden atribuir a cambios en la utilización del programa asociados estadísticamente con cambios en la variable instrumental. La intuición es que los efectos sobre los resultados económicos de estos cambios en la utilización del programa inducidos por el instrumento muestran los verdaderos efectos del programa, porque estos cambios en la utilización del programa no estarán correlacionados con variables no observables que predicen los resultados económicos, ya que estos cambios están generados por una variable instrumental no correlacionada con variables no observables que predicen los resultados económicos. El problema es encontrar este tipo de instrumentos. La variable instrumental debe dar una buena explicación de la utilización del programa, de lo contrario el enfoque de estimación elimina demasiada información. Pero la variable debe tener poca (idealmente cero) correlación con las variables no observables que afectan los resultados económicos, y es difícil verificar los supuestos sobre la correlación de un instrumento propuesto con variables no observables. Los buenos instrumentos son difíciles de encontrar y pueden no ser convincentes para todos los lectores.

Estos cinco enfoques se pueden combinar de diferentes maneras. Por ejemplo, los investigadores pueden crear conjuntos de datos emparejados, incluyendo controles para distintas variables observadas en la estimación, y utilizar variables instrumentales para la variable del programa.

Hay muchos ejemplos buenos de evaluaciones del impacto resultante de políticas de desarrollo local que utilizan datos no experimentales. Para los programas que proporcionan apoyo a firmas seleccionadas, Holzer *et al.* (1993) utilizaron implícitamente un enfoque de variable instrumental para estudiar un programa que proporciona capacitación a sus trabajadores adecuada a las necesidades de la empresa. Este programa, operado por el estado de Michigan, proporciona subvenciones a las empresas de manufactura para el entrenamiento de los trabajadores. El grupo de comparación estaba compuesto por firmas que solicitaron la subvención demasiado tarde en el año fiscal como para recibirla. El supuesto implícito es que el momento en que la firma solicitó la subvención es un “instrumento” que explica la participación en el programa, pero no está correlacionado con variables no observables que afectan el desempeño de la empresa. El estudio halló que las firmas que recibieron las subvenciones tuvieron un menor índice de desperdicio después de terminada la capacitación que las firmas que las solicitaron pero no las recibieron. El estudio produciría resultados sesgados si las firmas que postularon tarde en el año fiscal diferían de maneras que no podemos controlar (*v. gr.*, si dichas firmas estaban peor administradas).

Otro buen estudio de un programa selectivo de empresas es la evaluación de Jarmin de la Asociación para la Ampliación de la Manufactura (MEP, por sus siglas en inglés), patrocinada federalmente, que proporciona consejería por consultoría a las pequeñas y medianas empresas de manufactura para mejorar su productividad (Jarmin 1999). En el artículo de investigación de Jarmin primero se hace un emparejamiento aproximado incluyendo solamente a no clientes en los datos si estaban ubicados en los dos estados donde estaban todos los clientes. A continuación, en el artículo se controla el sesgo de selección mediante la estimación de una ecuación que determina si una empresa de manufactura dada se vuelve cliente de la MEP (por ejemplo, esto queda afectado si la empresa está en una zona metropolitana que tiene un centro de la MEP), e incluyendo una corrección del sesgo de selección en dos ecuaciones que predicen el crecimiento de la productividad de la firma; una ecuación para las empresas que son clientes de la MEP y otra ecuación para empresas que no lo son. El estudio de Jarmin halló que la MEP incrementa la productividad entre un 3 y un 16 por ciento.

Para zonas objetivo, varios estudios en los Estados Unidos han intentado evaluar las zonas empresariales mediante la comparación del desempeño de tales zonas con zonas emparejadas que no son empresariales. Diversos estudios realizados por investigadores de la Universidad Carnegie-Mellon (Bondonio y Engberg 2000, Greenbaum 1998, Greenbaum y Engberg 1998), han efectuado explícitamente tales emparejamientos utilizando estimaciones de la "propensión a participar", es decir, estimaciones de la probabilidad de que una zona dada (en este caso, una "zona" postal o código de ruta) sea designada zona empresarial. Estos estudios hallaron poco o ningún efecto de la designación de zona empresarial sobre el crecimiento de la actividad empresarial local. Además, como ya se mencionó, la estimación de la propensión a participar indica que la designación de zona empresarial no está fuertemente correlacionada con el crecimiento previo del área, lo que incrementa las posibilidades de que las estimaciones revelen el verdadero efecto de la designación de zona empresarial. Otros estudios han analizado también el desempeño de las zonas empresariales con zonas que no lo son (*v. gr.*, Papke 1993, 1994; Hebert *et al.*, 2001), pero las versiones publicadas no contienen suficiente información para juzgar la validez del emparejamiento. Finalmente, un estudio de próxima publicación (Peters y Fisher 2002) evalúa las zonas empresariales designadas por el gobierno estatal comparando el desempeño de zonas empresariales con niveles de incentivos altos *versus* bajos. El supuesto es que los factores no observables que afectan el desempeño de una zona podrían estar correlacionados con la designación de la zona como empresarial, pero no están necesariamente correlacionados con la magnitud de los incentivos de la zona empresarial, los cuales dependen de los diversos compromisos políticos presentes en la propuesta de zona empresarial decretada por el estado en particular. Este estudio halló pocos efectos de la magnitud de los incentivos de la zona empresarial sobre el inicio o la expansión de empresas.

Para los programas de desarrollo económico local que asisten a todas las firmas elegibles en una zona determinada, existe una abundante bibliografía sobre los efectos de los impuestos estatales y municipales en la ubicación y crecimiento de los negocios que ha sido sintetizada por Bartik (1991, 1992) y Wasylenko (1997). Estos estudios normalmente abordan posibles correlaciones de los impuestos con otras variables que afectan la ubicación y el crecimiento de los negocios mediante la inclusión como variables explicativas de todos los factores pertinentes posibles para la ubicación y el crecimiento en las ecuaciones de estimación. Como lo resumen Bartik y Wasylenko, estos estudios generalmente reportan a una elasticidad de la actividad empresarial con respecto a los impuestos empresariales estatales y municipales en el rango de ¡0.1! a 0.6. Esto es, una reducción del 10 por ciento en los impuestos estatales y municipales empresariales llevará eventualmente a un incremento en la actividad empresarial del estado del 1 al 6 por ciento.

En los últimos diez años, Andrew Isserman y sus colegas han elaborado una serie de artículos que evalúan diversas intervenciones de desarrollo económico en los Estados Unidos emparejando los condados intervenidos con condados de comparación sin intervenciones, utilizando datos anteriores a la intervención. Uno de tales estudios indicaba que la Comisión Regional de los Apalaches (CRA) incrementó el crecimiento en los condados de los Apalaches en comparación con los condados fuera de los Apalaches emparejados, siendo este efecto de crecimiento mayor en los condados en los que la CRA construyó autopistas (Isserman y Rephann 1995). Otro estudio evaluaba un gran recorte impositivo en el estado de Illinois mediante el emparejamiento de cada condado en Illinois con condados similares fuera del estado, y halló que el recorte fiscal tuvo algunos efectos de crecimiento económico en el corto plazo pero no efectos significativos de largo plazo (Rogers y Reed, próxima publicación).

Por último, uno de los mejores estudios de los efectos de los impuestos estatales compara la decisión de ubicación empresarial, entre los estados de Estados Unidos de empresas extranjeras procedentes de dos grupos de países: países en los que los impuestos estatales pagados en Estados Unidos pueden ser acreditados frente a responsabilidades fiscales en sus países de origen, y países en los que los impuestos en los Estados Unidos sólo pueden ser deducidos frente a los ingresos acumulables en el país de origen (Hines 1996). Para las firmas de los primeros países, los impuestos estatales en los Estados Unidos no deberían ser pertinentes en la decisión de ubicación empresarial para cualquier firma con obligaciones fiscales positivas en su país de origen. Hines (1996) halló que las firmas del primer grupo de países se ubicaban en estados de Estados Unidos con impuestos más altos que las firmas del segundo grupo de países. Esto se puede considerar como una forma de estimación mediante la DD o por variable instrumental. El supuesto implícito es que la única diferencia entre los dos grupos

de empresas que es pertinente para la decisión de ubicación empresarial es cómo el país de origen trata los impuestos estatales en Estados Unidos. Si este supuesto se mantiene, entonces las estimaciones resultantes son evidencia convincente de que los impuestos estatales y municipales afectan la decisión de ubicación empresarial de las grandes corporaciones en los Estados Unidos.

Si una localidad tiene un enfoque de desarrollo económico verdaderamente “único”, ¿se puede evaluar?

Los experimentos o los análisis estadísticos que utilizan grupos de comparación suponen que se poseen datos sobre un número significativo de firmas o zonas que utilizan el programa. Estos métodos estadísticos requieren un tamaño de muestra suficiente de usuarios del programa de modo que se pueda suponer que los factores únicos que afectan los resultados económicos de los usuarios se promedian en la muestra. ¿Pero qué sucede si una zona tiene un programa de desarrollo económico único? Por ejemplo, ¿qué sucede si una zona tiene algún paquete único de programas de desarrollo económico que se cree que tienen efectos de sinergia, de modo que el efecto completo del paquete no se puede predecir con precisión aun cuando se conocen los efectos sobre los resultados económicos de cada programa en lo individual?⁶

Si este paquete de programas de desarrollo económico se ofrece a un grupo selecto de firmas elegibles, entonces las firmas que no usen el programa se pueden utilizar como programa de control o de comparación, ya sea utilizando métodos experimentales o análisis estadísticos no experimentales, dependiendo de la disponibilidad de recursos y de voluntad de evaluación. Los procedimientos serían idénticos a los que previamente se han descrito. Asimismo, si el paquete se ofrece a un grupo objetivo de zonas marginadas, entonces las zonas marginadas que no son objetivos se pueden utilizar como grupo de control o de comparación, como se ha descrito anteriormente.

Por otro lado, ¿qué sucede si el paquete se ofrece a todas las firmas elegibles en una zona completa? En este caso, lo mejor que cualquier análisis estadístico puede determinar es si los resultados económicos de la zona difieren significativamente del desempeño promedio que uno predeciría, con base en el desempeño de zonas de comparación emparejadas o con base en una ecuación de predicción que utilice las características de las zonas de comparación. Si sólo una zona ofrece este programa, sus resultados económicos tendrán que diferir bastante más del promedio predicho para otras zonas para que sus resultados sean estadísticamente diferentes de manera significativa, en comparación con una situación en la que un número considerable de zonas ofrecen el mismo programa. Además, incluso si el desempeño de la zona es estadísticamente diferente de manera significativa de lo que se hubiera predicho, todo lo que se

puede concluir es que los efectos netos de los programas de desarrollo económico ampliados únicamente en la zona, y cualquier otra característica especial de la zona durante este periodo de tiempo, resultaban en un efecto neto sobre los resultados económicos significativamente diferente. Separar claramente lo que es atribuible al programa único es imposible.

¿Hay otras evidencias aparte de las comparaciones estadísticas con grupos de control o de comparación que puedan indicar el impacto del programa?

Dadas las dificultades y la incertidumbre asociadas con los experimentos o comparaciones estadísticas de los programas, es importante considerar si existen métodos alternativos que puedan sustituir o complementar los experimentos o las comparaciones estadísticas, y que nos permitan inferir algún vínculo entre las actividades del programa de desarrollo económico y los resultados económicos. Yo sostengo que hay al menos tres métodos alternativos para vincular las actividades del programa y los resultados económicos. Primero, en algunos casos, si uno cree que se puede modelar de qué manera los diferentes programas afectan las decisiones empresariales, entonces puede que sea posible extrapolar los resultados obtenidos en otros programas a nuevos programas. Por ejemplo, con las actuales tasas de impuestos estatales y municipales a los negocios en Estados Unidos, una elasticidad de la actividad empresarial estatal con respecto a los impuestos empresariales estatales/municipales del 0.25, próxima a la mediana del resultado en los estudios de investigación, implica que la creación de un puesto de trabajo cuesta aproximadamente 9000 dólares anuales en ingresos fiscales a las empresas en los estados y municipios o, utilizando una tasa de descuento real del 10 por ciento, un valor presente de los ingresos fiscales a empresas en los estados y municipios de 90000 dólares (Bartik, Eisinger y Erickcek 2003; Bartik 1992). Si se está dispuesto a suponer que lo que importa en un subsidio al desarrollo económico es su costo y que todas las reducciones de costos tienen aproximadamente los mismos efectos en la probabilidad de ubicación empresarial, entonces se puede inferir un probable efecto de diferentes subsidios empresariales sobre el desarrollo económico. Por ejemplo, supongamos que se otorgan subsidios para el desarrollo económico a una planta de nueva división con 1000 empleados por un valor actual de 30 millones de dólares. Entonces para ser completamente congruentes con la bibliografía de impuestos y ubicación empresarial, sería de esperar que estos subsidios incrementaran las posibilidades de que la planta de división escogiera el estado en una tercera parte, porque este efecto en la probabilidad de ubicación produciría un valor presente del costo por empleo creado de \$90000 dólares ($\$30 \text{ millones} / [(1/3)(1000)]$). Por supuesto, en cualquier caso particular, el subsidio fue o no decisivo para inclinar la balanza hacia la decisión de ubicación. Si se posee información de que el subsidio fue decisivo en mayor o menor grado, se debería

utilizar. Pero a falta de otra información, no queda claro porqué los efectos de los subsidios para el desarrollo económico sobre las probabilidades de ubicación empresarial tendrían que diferir de los efectos de los impuestos empresariales estatales y municipales en general.⁷

Segundo, las encuestas a las empresas que reciben el apoyo de programas de desarrollo económico pueden, si se aplican correctamente, utilizarse para obtener una idea aproximada de los efectos de algunos programas de desarrollo económico sobre las decisiones empresariales. Se ha vuelto muy común para las agencias de desarrollo económico estatales y municipales la utilización de alguna clase de encuesta de “satisfacción del cliente” con los usuarios. Sin embargo, las encuestas más útiles hacen preguntas específicas sobre cómo ha afectado su comportamiento el apoyo provisto a la empresa. Las respuestas a tales preguntas en las encuestas de “impacto resultante” son más creíbles cuando se pregunta por un apoyo al desarrollo económico entregado mediante servicios en especie que en los casos de apoyo en efectivo, porque las empresas están incentivadas para aseverar que los apoyos en efectivo tuvieron impacto en mantener el flujo de fondos, mientras que no queda claro por qué una empresa aseveraría que un servicio en especie fue útil si en realidad fue inútil. Un buen ejemplo de encuestas de desarrollo económico para determinar el impacto de los servicios de desarrollo económico son las encuestas regulares de los clientes del programa de la Asociación para la Ampliación de la Manufactura (MEP por sus siglas en inglés), cuyos centros locales proporcionan apoyo a pequeños y medianos manufactureros para mejorar su productividad. En las encuestas más recientes de la MEP llevadas a cabo en 2001 para los clientes que concluyeron su proyecto el año anterior, cerca del 64 por ciento de los encuestados reportaron que su participación con la MEP los condujo a mejoras en la productividad (NIST 2002). El cliente promedio de la MEP reportó que sus servicios le llevaron a incrementos en las ventas de 143 000 dólares y a ahorros en costos de 50 000 dólares. No queda claro por qué los clientes de la MEP buscarían fabricar este tipo de respuestas, en particular porque se las proporcionaron anónimamente a una tercera organización de encuestas.

La credibilidad de las respuestas de las empresas a encuestas sobre el impacto de beneficios fiscales y otros apoyos financieros es más cuestionable. Las empresas tienen claramente algunos incentivos para aseverar que el apoyo financiero afectó su ubicación o comportamiento de expansión con la esperanza de mantener operando el apoyo. En algunos casos, incluso puede que el programa haya requerido que la empresa firmara una estipulación señalando que el subsidio era esencial para la decisión de ubicación o de expansión con el propósito de recibir un subsidio. Sin embargo, ciertamente no siempre sucede que las firmas aseveren que el apoyo financiero fue crucial. Por ejemplo, la oficina de auditoría de la legislatura del estado de Colorado, en una auditoría al programa de la zona empresarial del estado, encuestó a 18 empresas que se habían ubicado

y expandido en zonas empresariales de Colorado, y encontró que 10 de estas empresas reportaron que los incentivos de la zona empresarial no tuvieron efecto en su decisión de ubicación y expansión (Hinckley y Hsu, 2000). Las encuestas sobre apoyo financiero son más creíbles cuando las administra anónimamente una agencia independiente, sobre todo cuando las encuestas hacen preguntas específicas y definitivas (por ejemplo, “¿consideraron otras ubicaciones?” o “¿la ubicación que ustedes eligieron habría sido claramente inferior en rentabilidad a estas otras ubicaciones sin el subsidio?”).

Un tercer método para inferir una conexión entre las actividades del programa y los resultados económicos locales es determinar si los datos administrativos sobre el programa y sus clientes son congruentes con el propósito asentado para el programa. Por ejemplo, el Programa de Acceso al Capital (CAP, por sus siglas en inglés) en Michigan fue diseñado para alentar a los bancos a proporcionar préstamos de alto riesgo a pequeñas empresas prestatarias. Por cada programa de préstamos a pequeñas empresas, el banco y el prestatario depositaban cada uno de 1.5 a 3.5 por ciento del valor del préstamo en un fondo de reserva para pérdidas de préstamos, y el CAP proporcionaba un tanto del 150 por ciento de la contribución del banco y los prestatarios al fondo. Los datos administrativos indican que el programa resultante tuvo un índice de pérdidas de cerca de siete veces el índice de pérdidas normales de un banco sobre préstamos a pequeñas empresas, lo cual significa que, como mínimo, el programa está alentando probablemente préstamos que de otra manera no se habrían concedido (Rohde, Cash y Ammerman 1990). Este hallazgo es congruente con la hipótesis de que el programa expande la oferta de crédito a las pequeñas empresas, sin embargo no prueba definitivamente que el programa pasaría una prueba de costo-beneficio. Como otro ejemplo, el programa MEGA en Michigan proporciona créditos fiscales reembolsables muy grandes a un grupo selecto de nuevas empresas o de empresas en expansión, pero sólo si la empresa puede presentar información financiera que demuestre que sin el subsidio la firma se ubicaría fuera del estado. Aunque es obviamente posible que las empresas maquillen este tipo de información financiera, el estado es libre de hacer preguntas para sondear el análisis de datos elaborado por la empresa y negarse a conceder créditos MEGA si las respuestas de la empresa no son suficientemente convincentes. El requisito de que la firma demuestre financieramente que el subsidio será decisivo al menos aumenta las dificultades y los costos a las empresas con razones relativamente débiles para solicitar el programa MEGA. Los funcionarios públicos en la agencia de desarrollo económico de Michigan afirman que excluyen a más del 90 por ciento de las firmas que expresan interés por el programa MEGA (Bartik, Eisinger y Erickcek, 2003).

¿Podemos determinar cómo y por qué un programa tiene impactos o no logra tenerlos?

Una preocupación sobre las evaluaciones del impacto resultante es que a menudo se perciben, incluso cuando están bien realizadas, como si sólo dijeran si un programa funciona o no, pero dejan los entresijos del programa en una “caja negra”: no sabemos porqué o cómo funciona el programa, o sea que no sabemos cómo mejorarlo. En principio, el análisis estadístico utilizando grupos de control o de comparación puede proporcionar percepciones de cómo y porqué funciona un programa, si se observa una variación suficiente en los diseños del programa y se mide con precisión. Con datos sobre muchas evaluaciones, la comparación estadística con grupos de control o de comparación puede indicar qué diseños de programa son más eficaces, lo cual en términos prácticos es tan positivo como saber cómo y porqué funciona un programa.

Sin embargo, en el mundo real, raras veces se observa la variación suficiente en los diseños de programa para responder adecuadamente a todas las preguntas importantes sobre cómo y porqué funciona el programa. En particular, en principio es imposible tener datos sobre las variaciones de un programa que todavía no se ha probado.

Lo anterior indica que las encuestas a clientes y los grupos focales de clientes pueden ser a menudo valiosos para abrir la “caja negra” y para proporcionar una mayor percepción de las fortalezas y debilidades del programa. El análisis estadístico que utiliza grupos de control o de comparación y las encuestas y los grupos focales se deberían considerar enfoques complementarios para evaluar un programa de desarrollo económico local. Es más probable que las comparaciones estadísticas proporcionen evidencia cuantitativa objetiva de los fundamentos del programa —su impacto en las condiciones económicas locales— que las encuestas y los grupos de enfoque, más aptos para proporcionar información sobre cómo se pueden mejorar esos fundamentos.

¿Podemos determinar los impactos de un programa sobre los objetivos finales y no sólo sobre los próximos?

Hasta ahora, en los debates no se ha considerado qué resultados económicos se deberían evaluar. En la práctica, los impactos económicos más fáciles de evaluar son los impactos próximos sobre varias dimensiones de la actividad empresarial, tales como la cantidad de aperturas o expansiones de empresas, el crecimiento del empleo, el crecimiento de la productividad, etcétera. Pero los subsidios públicos a los programas de desarrollo económico local no se pueden justificar únicamente por los efectos de estos programas en la actividad empresarial local. Los subsidios públicos al desarrollo económico local exigen que los cambios en la actividad empresarial local conduzcan a beneficios públicos más amplios. Los beneficios

públicos más plausibles son los beneficios fiscales para los gobiernos estatales y municipales, y los beneficios del empleo para los residentes locales.

Existen razones fundamentales de eficacia para algunos programas que promueven una mayor actividad empresarial, aun cuando no haya beneficios públicos más generales. Por ejemplo, algunas políticas de desarrollo económico local se pueden justificar por diversas “fallas de mercado” en los mercados de información o en los financieros (Bartik, 1990). A veces, los mercados privados de información no logran proporcionar a las empresas información que tenga más valor que lo que pagan por ella, justificando el suministro o el subsidio público de la información. A veces, los mercados financieros no logran otorgar préstamos o hacer inversiones en empresas cuyos beneficios privados excedan los costos, justificando potencialmente algunos subsidios o inversiones públicas a préstamos o inversiones privadas. Pero si todo lo que hacen esas intervenciones públicas es promover mayor actividad empresarial, sin mayores beneficios públicos, no queda claro por qué el público en general ha de pagar por estas intervenciones; la comunidad empresarial en general sería una fuente más justificable de fondos.

El público se beneficia del desarrollo económico local si la mayor actividad empresarial local conduce a beneficios fiscales o de empleo (Bartik, 1991). Los beneficios fiscales se dan cuando la mayor actividad empresarial, y los efectos derivados de ella sobre la economía local, tienen por resultado ingresos fiscales que exceden los incrementos en el gasto público requeridos. Los beneficios en empleo se dan cuando los salarios de los puestos de trabajo de reciente creación exceden los “costos de oportunidad” del tiempo libre al que renuncian los residentes locales que consiguen trabajo gracias a los empleos recién creados. Los nuevos empleos deben ser cubiertos por residentes locales empleados, residentes locales desempleados y por inmigrantes. Los puestos de trabajo cubiertos por residentes locales empleados generan vacantes que son cubiertas de la misma manera, o sea que en definitiva los empleos recién creados son cubiertos por residentes locales desempleados o por inmigrantes. Si no se hubieran creado estos empleos, los inmigrantes se podrían haber trasladado a otra zona local similar y obtener un empleo, por lo que el costo de oportunidad de su tiempo es equivalente a su nivel salarial. Para los residentes locales desempleados, el costo de oportunidad de su tiempo (su “salario de reserva”) puede ser considerablemente menor que el nivel salarial de los nuevos empleos.

Para calcular los beneficios fiscales y en empleo derivados de las políticas de desarrollo económico local se requiere de un modelo económico que incorpore los efectos iniciales de las políticas sobre la actividad empresarial local y que calcule los impactos sobre la economía local en general, incluyendo los efectos multiplicadores en proveedores y vendedores al menudeo, y los efectos en el crecimiento de la población local. Comúnmente las agencias de desarrollo económico de los Estados Unidos utilizan una variedad de esos modelos

regionales, siendo los más prominentes el modelo REMI y el modelo IMPLAN. Una vez determinados los efectos generales sobre toda la actividad económica local y el crecimiento de la población, es necesario traducir estos impactos en impactos en los presupuestos gubernamentales estatales y municipales y en beneficios locales de empleo.⁸

Los modelos de impacto fiscal que traducen los impactos económicos en impactos presupuestales requieren ser construidos especialmente para cada zona estatal o municipal, dadas las variaciones en la economía local y las estructuras fiscales y presupuestales locales. Los impactos fiscales sobre los gobiernos estatales y municipales dependen de varios factores. Primero, los impactos fiscales dependen, en parte, de cómo fiscalizan las diferentes bases impositivas los gobiernos estatales y municipales. Segundo, los impactos fiscales dependen del monto de inmigración poblacional comparado con el incremento en la actividad empresarial, ya que en los Estados Unidos el caso general es que la empresa promedio pague más en impuestos normales de lo que requiere directamente en servicios públicos, mientras que el hogar promedio usa más servicios públicos de lo que paga en impuestos (Oakland y Testa 2000). Tercero, los impactos fiscales dependen, en parte, de la proporción del gasto estatal o municipal que se destina a propósitos relacionados con la redistribución del ingreso, tales como la asistencia social o el Medicaid (el programa de apoyo médico de los Estados Unidos para los pobres), ya que este tipo de gasto no responderá proporcionalmente al crecimiento económico local. Cuarto, los impactos fiscales dependen en parte de si la infraestructura estatal y municipal existente tiene aún capacidad no utilizada que permita la expansión económica sin requerir la costosa construcción de nueva infraestructura.

En los Estados Unidos, los impactos fiscales positivos del desarrollo económico local pueden ser a menudo significativos. Por ejemplo, un estudio reciente calcula que los impactos fiscales positivos del desarrollo económico estatal en Michigan pueden llegar a compensar la mitad de los costos brutos de los subsidios estatales para el desarrollo económico (Bartik, Eisinger y Erickcek, 2003).

Estudios de mercados laborales locales en los Estados Unidos indican que por cada uno por ciento de crecimiento del empleo extra en un área metropolitana, los índices de empleo local crecen un 0.2 por ciento, y los ingresos promedio por empleo crecen debido a la mejora ocupacional un 0.2, de modo que los ingresos promedio por residente local crecen cerca de un 0.4 por ciento (Bartik, 1991). Estos efectos en los ingresos pueden ser perdurables si la experiencia de empleo extra para residentes locales previamente empleados incrementa sus competencias laborales, la confianza en sí mismos y su reputación con los empleadores (Bartik 2001). Otras estimaciones sugieren que estos efectos positivos sobre los ingresos locales son mayores si el crecimiento extra del empleo se concentra en trabajos que pagan bien en relación con las habilidades requeridas, tales como

los trabajos de manufactura (Bartik, 1993). Los modelos teóricos de mercados laborales locales indican que los salarios de reserva de los residentes locales recién empleados son probablemente mayores en las áreas metropolitanas con índices de desempleo superiores, en las que es más probable que los desempleados estén urgidos de conseguir un empleo (Bartik, 1991). Algunas agencias estatales de desarrollo económico, muy notablemente en el estado de Nueva York (Poole et al., 1999), han hecho algunos intentos de aplicar estos modelos del mercado laboral local a la evaluación de políticas de desarrollo económico local. Sin embargo, sabemos menos de lo que deberíamos sobre cómo los beneficios en el empleo derivados de las políticas de desarrollo económico local son afectados por otros factores. Por ejemplo, sería de esperar que los beneficios en empleo sean mayores cuando existe un mejor empate entre las plazas recién creadas y las habilidades laborales de los desempleados locales, pero no existe evidencia empírica directa que respalde esta expectativa. También sería de esperar que los beneficios en empleo sean mayores cuando las instituciones locales del mercado laboral son más eficaces en la capacitación para el trabajo y en la adecuación con el trabajo, pero nuevamente no existe evidencia empírica directa. Más estudios de los factores que afectan la relación entre el desarrollo económico local y los beneficios en empleo ayudarían a mejorar la calidad de las evaluaciones de las políticas de desarrollo económico local.

Conclusión

Este artículo de investigación ha tratado de mostrar que es factible una evaluación más rigurosa de los programas de desarrollo económico. Dicha evaluación rigurosa se puede realizar mediante experimentos aleatorios, análisis estadísticos de los usuarios del programa y grupos de comparación, encuestas y grupos de enfoque, relacionando modelos econométricos regionales con los impactos fiscales y modelos del mercado laboral local.

En los Estados Unidos se han llevado a cabo ampliamente esta clase de evaluaciones rigurosas. Estos estudios a menudo encuentran que los servicios a empresas de manufactura de tamaño pequeño y mediano pueden ser efectivos para mejorar el desempeño de estas empresas. Los programas que se dirigen a zonas marginadas tienden a ser inefectivos si, como las zonas empresariales, proporcionan recursos modestos, pero son más fructíferos si, como la Comisión Regional de los Apalaches, proporcionan gran cantidad de recursos durante un periodo prolongado. Los programas que proporcionan incentivos financieros o servicios para alentar el desarrollo económico de toda una zona, como incentivos fiscales estatales, generan efectos modestos en el aumento del crecimiento del empleo, lo cual es razonable dada la modesta porción de los costos sobre la que pueden influir los gobiernos estatales o municipales. Sin embargo, si un estado o una ciudad tiene una extensa infraestructura o mano de obra subutilizada,

incluso modestos incrementos en el empleo pueden ofrecer beneficios fiscales y de empleo considerables.

No obstante, la evaluación rigurosa sigue siendo la excepción y no la regla. Hay demasiados casos en los que organizaciones de desarrollo económico estatales y municipales reclaman el crédito de cualquier crecimiento del empleo estatal o local, o al menos de cualquier crecimiento derivado de empresas que esas organizaciones han subsidiado.

¿Cómo puede alentarse una evaluación más rigurosa? Veo pocas oportunidades de un incremento significativo de evaluaciones rigurosas sin presión y financiamiento externos. Las organizaciones de desarrollo económico, al menos en los Estados Unidos, enfrentan una significativa falta de incentivos para evaluarse rigurosamente a sí mismas. En el contexto de Estados Unidos, con la sospecha de activismo por parte del gobierno, las evaluaciones negativas son una excusa común para dar por terminado un programa. Por ello, a los diseñadores de políticas les suele parecer sensato el consejo de los economistas Gary Burtless y Robert Haveman: “Cuando se aboga por una reforma o innovación de alguna política en particular, no hay que presionar para que se ponga a prueba” (Burtless y Haveman, 1984, p. 128).

Por lo tanto, no me sorprendió hace unos años que un funcionario público del desarrollo estatal me dijera que el problema de que las universidades evaluaran sus programas es que parecían pensar que las evaluaciones negativas debían hacerse públicas. Se puede deplorar esta actitud, pero hay que entender el miedo real a recortes presupuestales y a la extinción del programa que la motiva.

Una evaluación rigurosa de políticas de desarrollo económico sólo es probable que ocurra si la financian o la requieren grupos externos. Estos grupos externos podrían incluir a legislaturas, oficinas de auditoría gubernamentales y altos escalafones del gobierno. Para estos grupos tiene sentido requerir y pagar evaluaciones porque los beneficios de la evaluación se acumulan en gran parte fuera de la agencia que administra el programa. Las evaluaciones de programas de desarrollo económico de agencias estatales o municipales proporcionan beneficios al público en general de la zona local, que se beneficia de cualquier mejora que tenga por resultado la eficacia del gobierno, y a localidades en otras partes que tengan programas similares o pudieran tomarlos en consideración.

Es también más probable que haya evaluaciones rigurosas si los resultados se utilizan más a menudo para mejorar programas que para ponerles fin. Si la lógica básica del programa tiene sentido en cuanto a que el programa aborde un problema que puede beneficiarse de la intervención del gobierno, los resultados negativos de la evaluación se deberían utilizar para motivar la creación de un nuevo enfoque para abordar el problema. Esto es más probable si la evaluación rigurosa va acompañada de datos procedentes de encuestas y grupos focales

que proporcionen alguna percepción de cómo se puede mejorar el programa. Una mezcla equilibrada de técnicas de evaluación rigurosas y “blandas”, y un uso juicioso de los resultados de la evaluación alentarán a las agencias de desarrollo económico a mostrar más apertura a enfoques de evaluación rigurosos.

Notas

1. He considerado con anterioridad estos aspectos en Bartik y Bingham (1997). El presente artículo de investigación actualiza mi pensamiento sobre este tema y toma en consideración hallazgos de investigación más recientes.
2. Las fuentes de estas estimaciones se analizan en Bartik (2001, p. 251) y son congruentes con estimaciones más recientes en Bartik, Eisinger y Erickcek (2003). En los Estados Unidos, a diferencia de Europa, no hay una recolección sistemática de datos sobre los presupuestos y gastos fiscales de los programas de desarrollo económico estatales y municipales (Thomas 2000). Estas cifras se han de extrapolar de estudios estatales individuales.
3. Por ejemplo, Thomas (2000) obtiene una cifra próxima a los 50 mil millones de dólares anuales en subsidios empresariales estatales/municipales incluyendo algunos de los gastos fiscales estatales y municipales más generales para las empresas.
4. Esta exposición palia la cuestión de que el efecto del programa puede variar entre las empresas. Técnicamente, todo lo que este experimento puede estimar es el efecto del programa en el 25 por ciento adicional de todas las empresas inducidas por la mercadotecnia a utilizar el programa, el cual puede diferir del efecto de productividad en el 10 por ciento de las empresas que utilizan el programa sin la promoción, o del 65 por ciento remanente de las empresas no afectadas por la promoción. A esto se lo denomina en ocasiones “Efecto Local Promedio del Tratamiento” (LATE, por sus siglas en inglés) o “efecto marginal del tratamiento” (Heckman y Vytlacil, 2001, Imbens y Angrist 1994).
5. Gran parte de lo que se expone a continuación en este artículo está formulado como si el programa se midiera por una variable dicotómica cero-uno, en la que el grupo de tratamiento utiliza el programa y el grupo de comparación no. Sin embargo, la explicación es generalizable a una situación en la que haya diferentes niveles de utilización del programa y en la que la variable de utilización del programa sea una variable continua.
6. Si no hay ningún efecto de sinergia especial de programas múltiples, entonces los efectos de una combinación “única” de programas se puede extrapolar a partir de estudios de los efectos de cada programa individual.
7. Este supuesto de que lo que importa son los costos empresariales totales se basa en el supuesto de que los efectos de producción dominan en la determinación de la ubicación de la empresa y el comportamiento de expansión, y en que los efectos de sustitución de factor son de importancia secundaria.
8. Para una exposición más extensa de modelos econométricos estatales y modelos asociados de impacto fiscal y de impacto en el empleo, y su uso en la evaluación de los incentivos para el desarrollo económico, véase el informe de Poole, Erickcek, Iannone, McCrea y Salem, 1999.

Referencias bibliográficas

- ANGRIST, J.D. y A.B. KRUEGER (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", *Journal of Economic Perspectives* 15(4): 69-85.
- BARTIK, T.J. (1990), "The Market Failure Approach to Regional Economic Development Policy". *Economic Development Quarterly* 4(4): 361-370.
- BARTIK, T.J. (con P. EISINGER y G. ERICKCEK) (2003), "Economic Development Policy in Michigan". En *Michigan at the Millennium*, C. Ballard, P. Courant, D. Drake, R. Fisher y E. Gerber, comps. East Lansing, Michigan: Michigan State University Press.
- BARTIK, T.J. y R.D. BINGHAM (1997), "Can Economic Development Programs be Evaluated?" En *Dilemmas of Urban Economic Development*, R.D. Bingham y R. Mier, comps. (pp. 246-277). Thousand Oaks, California: Sage Publications, Inc.
- BENUS, J.M., M. WOOD y N. GROVER (1994), "A Comparative Analysis of the Washington and Massachusetts UI Self-Employment Demonstrations". Informe inédito preparado para el Departamento del Trabajo de los Estados Unidos, la Administración para el Empleo y la Capacitación, Servicio del Seguro de Desempleo bajo Contrato núm. 99-8-0803-98-047-01.
- BONDONIO, D., y J. ENGBERG (2000), "Enterprise Zones and Local Employment: Evidence from the States' Programs". *Regional Science and Urban Economics* 30: 519-549.
- BURTLESS, G., y R.H. HAVEMAN (1984), "Policy Lessons from Three Labor Market Experiments". En *Employment and Training R&D*, R. Thane Robson, comp. Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research, p. 128.
- CCC (1991), *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.
- CCC (1992), "The Effects of State and Local Taxes on Economic Development: A Review of Recent Research". *Economic Development Quarterly* 26: 102-110.
- CCC (1993), "Economic Development and Black Economic Success". Informe Técnico del Upjohn Institute núm. 93-001. Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.
- CCC (1994), "Tax Policy and Urban Development: Evidence from the Indiana Enterprise Program". *Journal of Public Economics* 54: 37-49.
- CCC (1995), *The Price of Federalism*. Washington, DC: The Brookings Institution.
- CCC (2001), *Jobs for the Poor: Can Labor Demand Policies Help?* Nueva York y Kalamazoo, Michigan: Russell Sage Foundation y W.E. Upjohn Institute for Employment Research.
- FISHER, P.S. y A.H. PETERS (1998), *Industrial Incentives: Competition Among American States and Cities*. Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.
- GREENBAUM, R.T. (1998), "An Evaluation of State Enterprise Zone Policies: Measuring the Impact on Business Decisions and Housing Market Outcomes". Tesis doctoral, Carnegie Mellon University, diciembre.

- GREENBAUM, R.T. y J.B. ENGBERG (1998), "The Impact of State Urban Enterprise Zones on Business Outcomes". Discussion paper núm. 98-20. Washington, DC: Center for Economic Studies, Bureau of the Census.
- HEBERT, S., A. VIDAL, G. MILLS, F. JAMES y D. GRUENSTEIN (2001), Interim Assessment of the Empowerment Zones and Enterprise Communities (EZ/EC) Program: A Progress Report. Washington, DC: Office of Policy Development and Research, US Department of Housing and Urban Development.
- HECKMAN, J.J., H. ICHIMURA y P. TODD (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme". *Review of Economic Studies* 64(4): 605-654.
- HECKMAN, J.J. y E. VYTLACIL (2001), "Policy-Relevant Treatment Effects". *The American Economic Review* 91(2): 105-111.
- HINES, J.R., Jr. (1996), "Altered States: Taxes and the Location of Foreign Direct Investment in America". *American Economic Review* 86(5): 1076-1094.
- HINKLEY, S., y F. HSU (2000), "Minding the Candy Store: State Audits of Economic Development". Washington, DC: *Good Jobs First*, septiembre.
- HOLZER, H.J., R.N. BLOCK, M. CHEATHAM y J.H. KNOTT (1993), "Are Training Subsidies For Firms Effective? The Michigan Experience". *Industrial and Labor Relations Review* 46 (July): 625-636.
- IMBENS, G. y J. ANGRIST (1994), "Identification and Estimation of Local Average Treatment Effects". *Econometrica* 62(2): 467-475.
- ISSERMAN, A. y T. REPHANN (1995), "The Economic Effects of the Appalachian Regional Commission: An Empirical Assessment of 26 Years of Regional Development Planning". *Journal of the American Planning Association* 61(3): 345-364.
- JARMIN, R.S. (1999), "Evaluating the Impact of Manufacturing Extension on Productivity Growth". *Journal of Policy Analysis and Management* 18(1): 99-119.
- MAZEROV, M. (2001), *The "Single Sales Factor" Formula for State Corporate Taxes: A Boon to Economic Development or a Costly Giveaway?* Washington, DC: Center on Budget and Policy Priorities.
- MCLURE, C.E., Jr. y W. HELLERSTEIN (2002), "Does Sales-Only Apportionment of Corporate Income Violate the Gatt?" NBER artículo de investigación núm. 9060. Cambridge, Massachusetts: National Bureau of Economic Research.
- MEYER, B.D. (1995), "Natural and Quasi-Experiments in Economics". *Journal of Business and Economic Statistics* 13(2): 151-161.
- MOLNAR, L.A., D.R. GRIVES, J. EDELSTEIN, R. DEPIETRO, H. SHERMAN, D. ADKINS y L. TORNATZKY (con Y. BATTS, G. FULTON y S. HAYHOW) (1997), *Impact of Incubator Investments*. Athens, Ohio y Ann Arbor, Michigan: NBIA Publications y The University of Michigan.
- MURNANE, R.J., S. NEWSTEAD y R.J. OLSEN (1985), "Comparing Public and Private Schools: The Puzzling Role of Selectivity Bias". *Journal of Business and Economic Statistics* 3:23-35.
- NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (2002), "The Manufacturing Extension Partnership: Delivering Measurable Returns to Its Clients". Washington, DC: US Department of Commerce, Technology Administration.

- NEXUS ASSOCIATES (1999), *A Record of Achievement: The Economic Impact of the Ben Franklin Partnership*, Belmont, Massachusetts: Nexus Associates, Inc.
- OAKLAND, W.H. y W.A. TESTA (2000), "The Benefit Principle as a Preferred Approach to Taxing Business in the Midwest". En *Economic Development Quarterly* 14(2): 154-164.
- PAPKE, L.E. (1993), "What Do We Know About Enterprise Zones". En *Tax Policy and Economy*, vol. 7, J.M. Poterba, comp. Cambridge, Massachusetts: MIT Press, pp. 37-72.
- PETERS, A.H. y P.S. FISHER (2002), *State Enterprise Zone Programs: Have They Worked?* Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.
- PETERSON, P.E. (1981), *City Limits*. Chicago: University of Chicago Press.
- POOLE, K.E., G.A. ERICKCEK, D.T. IANNONE, N. MCCREA y P. SALEM (1999), *Evaluating Business Development Incentives*. Washington, DC: National Association of State Development Agencies.
- REED, W.R. y C.L. ROGERS (2000), "Measurement Error and Endogeneity in Studies of State Tax Policy and Economic Growth". Artículo de investigación inédito, Departamento de Economía, University of Oklahoma, Norman, Oklahoma.
- ROGERS, C. y W.R. REED (próxima publicación), "Quasi-Experimental Control Group Methods for Estimating Policy Impacts". *Regional Science and Urban Economics*.
- ROHDE, S., J. CASH y K. AMMARMAN (1990), "Study of the Capital Access Program". Artículo de trabajo inédito, Michigan Strategic Fund, Lansing, Michigan.
- SMITH, J.A. y P.E. TODD (2001), "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods". *The American Economic Review* 91(2): 112-118.
- THOMAS, K.P. (2000), *Competing for Capital: Europe and North America in a Global Era*. Washington, DC: Georgetown University Press.
- WASYLENKO, M. (1997), "Taxation and Economic Development: The State of the Economic Literature". *New England Economic Review* (marzo/abril): 37B52.

Capítulo 5

Cuatro enfoques para mejorar las prácticas de evaluación en la Unión Europea: Un comentario sobre el artículo de Timothy Bartik “La evaluación del impacto de las políticas de desarrollo económico local en los resultados de la economía local: ¿qué se ha hecho y qué se puede hacer?”

Por
Daniele Bondonio,
Departamento de Política y Elección Pública,
Università del Piemonte Orientale,
Alessandria, Italia

Introducción

El propósito de este comentario es situar dentro del contexto de la Unión Europea (UE) las técnicas y métodos analizados y recomendados por Timothy Bartik en su artículo “La evaluación del impacto de políticas de desarrollo económico local en los resultados de la economía local: ¿qué se ha hecho y qué se puede hacer?” (Teniendo especial atención en los programas de desarrollo económico local/regional cofinanciados por los Fondos Estructurales de la UE). ¿Es posible afirmar, en este lado del Atlántico, que los programas de desarrollo económico local y regional patrocinados por la Unión Europea, han estado sujetos a rigurosas evaluaciones tal como asegura Timothy Bartik que ha ocurrido en muchos de los programas estadounidenses? Al utilizar el mismo marco metodológico del artículo de Timothy Bartik, puedo argumentar que, en este lado del Atlántico, es posible mejorar nuestras prácticas de evaluación en cuatro diferentes direcciones:

1. Ser más claros en lo que significa una evaluación rigurosa;
2. registrar mejores datos e información;
3. incorporar las necesidades de la evaluación al diseño de la política;
4. explotar la heterogeneidad de los diseños de implementación regional a lo largo de la UE.

Este comentario se basa en gran medida, de igual forma que el artículo de Timothy Bartik lo hace sobre la evaluación de programas de incentivo a las empresas, sobre mi experiencia en la evaluación de los programas de incentivos a la inversión para Pequeñas y Medianas Empresas (PYMEs) en las áreas Objetivo 2¹ (Ob.2) financiados por la UE y comparando estos incentivos con las Zonas de Empoderamiento Federales de los Estados Unidos. Según mi entender, las presentes y pasadas evaluaciones de los programas de incentivos para las PYMEs en áreas Ob.2, son en cierta medida, representativas de las prácticas de evaluación adoptadas para otras políticas de desarrollo económico local implementadas por la UE.

Ser más claros en lo que significa una evaluación rigurosa

Es necesario que comience esta sección con el planteamiento de una breve premisa. La terminología utilizada para evaluación para los programas de desarrollo económico local es ligeramente distinta entre la UE y los Estados Unidos (particularmente en aquella utilizada en las evaluaciones de los programas financiados conjuntamente por los Fondos Estructurales de la UE). Términos de uso generalizado en la UE como “*ex ante*”, “*in itinere*” y “*evaluación ex post*”, no se utilizan comúnmente en la bibliografía de evaluación de los Estados Unidos. Es por ello que lo que para la UE es “*evaluación ex ante*”, en los Estados Unidos es “*evaluación de la factibilidad del programa*” y es considerado como una parte del amplio conjunto de actividades de programación enfocadas al efectivo diseño de políticas. Asimismo, “*evaluación in itinere*” en los Estados Unidos equivale a “*monitoreo de actividades de programas*”. Debo hacer notar, que el artículo de Timothy Bartik se enfoca mayormente a lo que en la UE se refiere como “*evaluación ex post*” (que quizás incluya algunos métodos de entrevistas o grupos de enfoque que en la UE también se les darían el nombre de “*evaluación in itinere*”). En concordancia con la terminología utilizada por Timothy Bartik, adoptaré el término “*evaluación*” para referirme a la “*evaluación ex post*”.

Tal como lo presenta claramente Timothy Bartik en su artículo, son necesarios tres enfoques complementarios para una rigurosa evaluación de políticas para el desarrollo económico local que genere incentivos para el desarrollo empresarial:

1. Evaluación del impacto en resultados en dimensiones cercanas a la actividad empresarial;
2. estimación de los beneficios fiscales y de empleo en la economía local mediante un modelo regional (Como REMI² o Implan³);
3. realización de encuestas y grupos focales a clientes, a fin de mejorar la efectividad administrativa del programa.

La evaluación de impacto de resultados en dimensiones cercanas a la actividad empresarial (como empleo o gastos de inversión) [enfoque a)] es un componente crucial para la evaluación rigurosa. Este componente genera evidencia crucial acerca de la efectividad aproximada del programa, ya que hace una estimación de la diferencia entre “lo que pasó” y “lo que hubiera pasado” si no se hubieran dado las políticas. Estas estimaciones de impacto, pueden servir a su vez como base para estimar los beneficios fiscales y de empleo generados por el programa en la economía local a través de modelos macroeconómicos regionales que puedan permitir una valuación de costo-efecto de la intervención [(enfoque b)]. Por otro lado, las entrevistas y grupos focales de clientes pueden realizarse de manera independiente tanto para los enfoques a) y b) y constituir

“evaluación *in itinere*”; según la terminología utilizada en las evaluaciones de los Fondos Estructurales de la UE.

En lugar de rigurosas evaluaciones sobre los resultados y sus efectos, las prácticas actuales de evaluación de programas co-financiados por los Fondos Estructurales de UE, generalmente buscan medir solo “lo que ocurrió” en las áreas objetivo (o empresas), en lugar de estimar las diferencias entre “lo que ocurrió” y “lo que hubiera ocurrido de no haber sido por” las políticas. Conocer “lo que ocurrió”, si está cuantificado de manera precisa, puede ser un elemento útil debido a que genera información importante sobre la actividad del programa, misma que puede ser utilizada para mejorar su administración. Sin embargo, dado que nos estamos basando en el artículo de Timothy Bartik, es necesario aclarar que las medidas de “lo que ocurrió” no son suficientes para una evaluación rigurosa.

Ejemplos de dichas prácticas, pueden encontrarse en el informe de evaluación temática de los efectos los Fondos Estructurales en las PYMEs encargado por la Comisión Europea (Ernst y Young 1999) y dos informes que resumen las evaluaciones de la Comisión Europea de los programas enfocados en las áreas Ob.2 de 1989-93 (Malan 1998, Bachtler y Taylor 1999). En los informes Ernst y Young (1999) y Malan (1998), los cálculos del efecto de los programas de las regiones Ob.2 UE se conforman comparando el crecimiento económico de las regiones Ob.2 con el crecimiento económico de las regiones no incluidas en regiones Ob.2 en la UE. Sin embargo, tal procedimiento puede únicamente identificar el efecto actual de los programas Ob.2 sólo si las áreas asistidas y no asistidas tuvieran un desempeño similar en ausencia de la intervención. Por ello, dar todo el crédito de una mejora en el desempeño de las áreas Ob.2 al programa, no constituye una práctica rigurosa de evaluación, a menos que se incluyan en el análisis procedimientos estadísticos y econométricos fidedignos (por ejemplo. Bartik 1991, Bartik y Bigham 1995, Bondonio 2000, Bondonio 2002, Manski 1995, Moffit 1991, Smith 2000) que permitan verificar y separar las diferencias en el desempeño debido a diferentes características previas a la intervención entre las distintas áreas.

En algunos informes de evaluación encargados por administraciones regionales italianas para evaluar las medidas de incentivo empresarial implementados en regiones italianas Ob.2 [por ejemplo Ecoter “Docup Ab.2: Rapporto di valuatazione finale” (Docup Ob.2: Informe de evaluación final)], los estimados sobre el impacto del programa se determinaron sumando el número de empleos que serían creados por cada empresario cuando fuera completado el apoyo de inversión a cada uno. Estos datos fueron extraídos de las aplicaciones de los empresarios al momento de solicitar el apoyo.

En otras prácticas de evaluación se han estimado los efectos de los programas de asistencia con dinero en efectivo para PYMEs en áreas Ob.2, a través del cálculo del número total de empleos que empresarios entrevistados indicaron

como trabajos que no se hubieran creado de no haber recibido asistencia del programa (por ejemplo Malan 1998 y Ernst y Young 1999). Así como argumenta Timothy Bartik, este enfoque puede estar sesgado por la tendencia de los empresarios que contaron con asistencia a asegurar que los beneficios de dinero en efectivo tienen alto impacto, ya que de esta manera aseguran el mantenimiento futuro del otorgamiento de dinero en efectivo.

Debe de indicarse que al aplicar programas REMI, Implan o cualquier modelo regional macroeconómico para determinar los beneficios fiscales y de empleo en la economía local, ello no constituye una evaluación rigurosa si dichos modelos se basan en información no confiable sobre los resultados del impacto del programa en dimensiones aproximadas a la actividad empresarial. Al no haber evidencia confiable proveniente de una evaluación rigurosa de los resultados, las estimaciones del impacto de la economía local en su conjunto serían incorrectas al estar basadas en datos erróneos dados a los modelos macroeconómicos regionales. Por ejemplo, en el informe final de evaluación preparado por Ecoter para la región Piedmont en 1999, la estimación del impacto en el empleo en la economía local de las áreas Ob.2, fue determinado aplicando un conjunto de multiplicadores de un modelo macroeconómico regional directamente a la cantidad total de todas las inversiones aprobadas en el área (un procedimiento que de manera implícita asume que no hubiera habido ninguna inversión en el área en la ausencia del programa).

Registrar mejores datos

La mayoría de los informes de evaluación de los programas de incentivos a las empresas dirigidos a regiones geográficas específicas financiados por los fondos estructurales de la UE, enfatiza la necesidad de obtener mejores datos acerca de las actividades de los programas a fin de mejorar los procedimientos de evaluación. Mejorar los sistemas de monitoreo del programa es comúnmente la solución recomendada para obtener datos más precisos y confiables de los beneficiarios de los programas y en las cantidades de asistencia de dinero en efectivo o de los servicios que se les ofrecen (por ejemplo Ernst y Young, 1999).

Sin embargo, los métodos estadísticos correctos a menudo sacan ventaja de las diferencias de los resultados entre las empresas (o áreas) que reciben asistencia y aquellas que no la reciben, tanto antes como después de la intervención del programa (por ejemplo Bartik, 1991, Bartik y Bigham, 1995, Moffit 1991, Smith, 2000). Por lo que hay que estar concientes que únicamente tener información correcta y confiable sobre las áreas que cuentan con asistencia, no es suficiente para conducir una evaluación rigurosa. Se necesitaría contar con datos confiables y correctos tanto para las áreas objetivo como para las que no lo son, para utilizar de manera apropiada los procedimientos estadísticos.

En el caso de los programas de incentivos a las empresas focalizados a nivel regional, la evaluación rigurosa se beneficiará de los datos recolectados a nivel de planta por los sistemas estadísticos nacionales y europeos, los que contraponen la información de empleo con los datos socioeconómicos de los residentes de unidades geográficas estadísticas pequeñas.

En la UE, las NUTS_3s⁴ constituyen en la actualidad las unidades geográficas más pequeñas. Allí se encuentran a disposición los sistemas estadísticos oficiales, que cuentan con una confiabilidad muy buena para realizar comparaciones a través del tiempo. Sin embargo, las NUTS_3s no son lo suficientemente pequeñas como para permitir reconstruir las fronteras relevantes de las áreas objetivo (como las áreas Ob.2).

En Greenbaum y Bondonio (2003), fueron analizadas y comparadas las características de las áreas asistidas por los programas federales de las “Zonas de Empoderamiento” (ZE) de Estados Unidos con las áreas Ob.2 de la UE. En los Estados Unidos, la información que de manera precisa emparejaba las fronteras de las ZE, pudo ser recuperada combinando información de distintos censos (unidades geográficas estándar utilizadas por el Buró de Censos de los Estados Unidos). Sin embargo, para el caso de la UE no se pudo obtener información precisa de las características de las áreas Ob.2. Sólo con información confiable de unidades geográficas estandarizadas pequeñas (Como NUTS_5) hubiera sido posible reconstruir las características de las áreas Ob.2. Por otro lado, las NUTS_5 son muy difíciles de obtener y no permiten realizar comparaciones confiables a través del tiempo debido a que se basan en fronteras administrativas de ciudades, las que frecuentemente cambian con el tiempo.

Así como la existencia de un panel de datos confiable es importante para un análisis riguroso, las prácticas de evaluación en la UE, mejorarían en la medida que se construyan sistemas estadísticos integrados que permitan el fácil acceso a la información, la cual estaría ordenada en pequeñas unidades geográficas que permanecerían constantes a lo largo del tiempo.

Los sistemas integrados de datos en la UE, deberían también incluir registros de las empresas que reciban apoyo de todas las fuentes de asistencia pública agrupadas por los países de la UE o por regiones. La creación de tales registros es necesaria para comparar a las empresas asistidas con las que efectivamente no lo fueron, ya que se podría comparar con empresas que recibieron algún otro apoyo distinto al ofrecido por la UE. Las reglas mínimas que imponen un techo al total de asistencia pública admisible por empresas de la UE, lentamente está forzando a las administraciones de países individuales de la UE a llevar registros de todas las empresas subsidiadas. Integrar tales registros regionales en un archivo unificado europeo, sería de gran ayuda para generar evaluaciones más rigurosas a lo largo de la UE

Incorporar las necesidades de evaluación en el diseño de la política

Tal y como aparece en el artículo de Timothy Bartik, puede ser posible obtener información confiable para una evaluación rigurosa incorporando, en el diseño del programa para el desarrollo económico local, algunas de las necesidades que se señalen en la evaluación. Ello debería de ser considerado seriamente por la UE debido a que la implementación de cambios en los sistemas estadísticos europeos puede alargarse y resultar muy costosa. Una manera de datos confiables para una evaluación rigurosa, consistiría en designar a las áreas objetivo aquellas regiones cuyas fronteras se traslapan con las unidades geográficas existentes en los sistemas estadísticos europeos.

Otra forma de obtener datos confiables para una evaluación rigurosa, sería implementar políticas con protocolos de experimentación. Sin embargo, frecuentemente la fuerte oposición política por excluir áreas con mucha necesidad o empresas de asistencia social, socava la implementación de protocolos de experimentación a gran escala dirigidos a programas para el desarrollo económico local. Aún así, alguna forma de experimentación podría ser aceptable y la UE debería considerarlo seriamente, de manera particular el procedimiento propuesto por Timothy Bartik de seleccionar de manera aleatoria a las empresas para un mercadeo del programa dirigido a un grupo objetivo. Tal experimento, puede producir diferencias significativas en la tasa de utilización del programa entre el grupo de empresas atendidas (aquellas que reciben esfuerzo de mercadeo) y el grupo de control (las que no reciben esfuerzo de mercadeo). Las diferencias de la tasa de uso entre el grupo de empresas que reciben atención y aquellas del grupo de control de empresas elegibles pueden ser utilizadas para estimar de manera confiable el impacto del programa. Como este esquema no implica la exclusión arbitraria de cualquier empresa elegible de la intervención del programa, la implementación debería enfrentar una resistencia política mínima.

Explotar la heterogeneidad de los diseños de implementación regional a lo largo de la Unión Europea

Tal y como Timothy Bartik menciona en su artículo, el análisis estadístico que utiliza grupos de control, puede generar ideas sobre la manera en que un programa funciona, así como del porqué de su funcionamiento, siempre y cuando exista suficiente variación en los diseños de los programas y sean cuantificados de manera precisa.

Los programas patrocinados por la UE, cuentan con mucha variación en el diseño de la implementación de políticas a lo largo de las distintas regiones donde los programas Ob.2 se llevan a cabo. Esta heterogeneidad en los diseños de política (para estos y otros programas patrocinados por la UE), no debería considerarse como una amenaza a la validez del análisis porque limita la

posibilidad de comparar los resultados de evaluación a lo largo de la UE. Debería, por el contrario, concebirse como una excelente oportunidad para examinar la efectividad de una variedad de diseños de política y de las diferencias en la generosidad de los incentivos en efectivo o los servicios que se ofrecen a las empresas a las que se asiste.

Es necesario aprovechar la heterogeneidad. Para ello se deberían implementar métodos estadísticos apropiados para que las variaciones en las características de políticas de cada región puedan volverse operacionales, y que las tendencias económicas específicas de cada región o país sean controladas y conservadas por separado de las estimaciones del impacto de las características de políticas específicas para cada región (por ejemplo Bondonio, 2002). Además, si la disponibilidad de datos a nivel de planta lo permite, es importante resaltar que el análisis se puede implementar separando los resultados observados de las empresas (por ejemplo, crecimiento del empleo) en tres componentes:

- Cambios atribuidos a nuevas empresas atraídas a áreas donde se provee asistencia;
- cambios en empresas beneficiadas;
- cambios de empresas que interrumpen su actividad.

Ordenar los resultados de esta manera puede ser muy útil dado que permite la investigación sobre si ciertas características de la política son más efectivas para atraer nuevas empresas, en lugar de frenar la disminución de la producción ya existente (o viceversa). Incentivos que parecieran ser apropiados para atraer nuevos negocios podrían ser recomendados para áreas objetivo tales como parques industriales recientemente desarrollados, en vez de lugares donde los principales objetivos de inversión son empresas ya operando en áreas que reciben asistencia.

Notas

1. Regiones con producción industrial en disminución elegibles para asistencia patrocinada por la UE:
2. Modelos Económicos Regionales, Inc. (REMI ®), www.remi.com.
3. Grupo IMPLAN, www.implan.com.
4. NUTS (por sus siglas en inglés) significa Nomenclatura para Unidades para Estadísticas Territoriales que es la estructura de la jerarquía regional de cinco niveles utilizada para estandarizar los territorios económicos de la UE. Las áreas NUTS_3 están al centro de la estructura jerárquica y están conformadas por un grupo de unidades geográficas compuesto de jurisdicciones subnacionales del segundo nivel (comparable en muchos aspectos con los condados en los Estados Unidos). Las áreas NUTS_1 (la cuales son las unidades más grandes de la estructura jerárquica compuestas como grupos de regiones aledañas o estados correspondientes a las jurisdicciones subnacionales para cada nación de la UE) y las áreas NUTS_5 (compuestas como el grupo de jurisdicciones de ciudades o pueblos de los países de la UE) completan la jerarquía.

Referencias bibliográficas

- Bachtler, J. y S. Taylor (1999), "Objective 2: Experiences, Lessons, and Policy Implications", Final Report European Policies Research Centre, http://europa.eu.int/comm/regional_policy/sources/docgener/evaluation/pdf/finalrep_full.pdf.
- BARTIK, T.J. (1991), *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI:W.E. Upjohn Institute for Employment Research.
- BARTIK, T.J., R. BINGHAM (1995), *Can Economic Development Programs be Evaluated*, W.E. Upjohn Institute for Employment Research, Kalamazoo, MI:Staff Working Paper 95-29.
- BONDONIO, D. (2002), "Evaluating Decentralized Policies: A Method to Compare the Performance of Economic Development Programmes across Different Regions or States", *Evaluation*, vol. 8, núm. 1, pp. 101-124, 2002.
- GREENBAUM, R. y D. BONDONIO (2003), "A Comparative Evaluation of Spatially Targeted Economic Revitalization Programs in the European Union and the United States", ICER-International Center for Economic Research Working Paper Series núm. 3/03, enero.
- ERNST y YOUNG (1999) "Thematic Evaluation of Structural Fund Impacts on SMEs", *Synthesis Report*, European Commission.
- MALAN, J. (1998), "Translating Theory into Practice: Lessons from the Ex Post Evaluation of the 1989-1993 Objective 2 Programmes", informe presentado en la Seville Conference on Evaluation Practice in the Field of Structural Policies de 1977, http://europa.eu.int/comm/regional_policy/sources/docconf/seville/sevil:en.htm.
- MANSKI, C.F. (1995), *Identification Problems in the Social Sciences*, Cambridge, MA and London, UK: Harvard University Press.
- MOFFIT, R. (1991), "Program Evaluation with Nonexperimental Data", *Evaluation Review* 15, 3:291-314.
- SMITH, J. (2000), "A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies", *Swiss Journal of Economics and Statistics* 136(3): 1-22.

Capítulo 6

La evaluación de programas dirigidos al desarrollo local y regional: metodología y veinte años de experiencia utilizando el modelo “REMI Policy Insight”¹

Por
Dr. Frederick Treyz
y
Dr. George I. Treyz

Prefacio

Los diseñadores de políticas requieren evaluar el efecto total de los programas locales y regionales con el propósito de tomar decisiones informadas. Las propuestas de desarrollo tienen implicaciones económicas, sociales y demográficas que van más allá de sus efectos directos. Para entender estos efectos, los analistas necesitan utilizar un modelo exhaustivo económico de pronósticos y simulación. Los modelos de análisis de políticas locales y regionales muestran los efectos completos de los cambios que inducen las políticas en la economía local, incluyendo las consecuencias socioeconómicas que de otra manera podrían ser imprevistas o pasar inadvertidas.

Este artículo de investigación describe el modelo REMI para la Percepción de Políticas, modelo líder para el pronóstico de la economía regional y el análisis de las políticas. Más de cien institutos, universidades, agencias de gobierno y otras organizaciones utilizan modelos REMI específicos construidos a la medida para estados, distritos, ciudades y otras regiones. Los usuarios de estos modelos están localizados principalmente en los Estados Unidos (EU), pero también incluyen a organizaciones que utilizan o consideran la utilización de modelos para regiones de Bélgica, Francia, Alemania, Italia, los Países Bajos, España y el Reino Unido. La Comisión Europea ha contratado a REMI para que desarrolle modelos para la evaluación de inversiones de fondos para estructuras.

Los analistas utilizan el modelo REMI para evaluar los efectos económicos de programas de desarrollo económico, de inversiones en infraestructura para el transporte, de regulaciones ambientales y energéticas, y de otras políticas públicas que repercuten en la economía regional o local. Los estudios REMI incluyen evaluaciones de trenes de alta velocidad, nuevas autopistas, programas de incentivos fiscales para empresas, asuntos de recursos acuíferos, controles de la contaminación atmosférica, desregulación del servicio eléctrico y cientos de otras aplicaciones. Los usuarios suelen incorporar los análisis REMI a sus procesos para evaluar, por ejemplo, todas las propuestas de reubicación potencial de empresas en un determinado estado o ciudad.

REMI proporciona un sistema para el trabajo de modelación que muestra los efectos totales de la política, incluso aquellos no anticipados. Por ejemplo, una

política para reducir la contaminación atmosférica puede tener consecuencias en los costos de empresas que verán reducida su competitividad, pero la política también puede tener el efecto no pecuniario de hacer de la zona un lugar más placentero para vivir. En este caso, la pérdida de competitividad reducirá la producción total, pero un aire más puro fomentará la migración hacia la zona, lo cual aumentará la mano de obra. A su vez, este aumento incrementará la productividad laboral y reducirá los salarios, disminuyendo por lo tanto los costos y aumentando la competitividad, lo cual incrementará la actividad económica. El efecto neto de estas políticas sólo lo puede captar un modelo exhaustivo que abarque la migración económica, la fuerza de trabajo, la determinación de los salarios, los precios del suelo y el efecto de ubicación de la competitividad sobre el empleo y la producción total en la zona.

Otro ejemplo de una consecuencia no intencional que requiere un modelo es la construcción de un sistema de ferrocarril con paradas limitadas. Esto puede aumentar la competitividad en la venta de servicios entre las zonas con estaciones, pero también puede disminuir la competitividad de todos los servicios en las zonas sin estaciones. El resultado neto de estos cambios tiene ramificaciones que podrían ser imprevistas en zonas e industrias en particular y en la economía regional en su conjunto. Sin un modelo apropiado, sería imposible predecir la dirección de los cambios en la actividad económica en cada una de las áreas del modelo, tanto en lo que se refiere al empleo en industrias específicas como en la economía en general, cuando se combinan las mejoras en el transporte y los costos de estas mejoras.

Ya que el objetivo final de las políticas de desarrollo económico es mejorar las condiciones económicas y sociales, es importante predecir y evaluar los efectos totales de estas políticas sistemáticamente. Esto ayudará a seleccionar el conjunto de políticas que alcanzarán el máximo beneficio para el gasto propuesto. Para hacer posible lo anterior, consideramos que el modelo cuantitativo tiene que incorporar todas las interacciones clave en la economía y basarse en una teoría económica bien fundada. Este tipo de modelo también se ha de diseñar para utilizar relaciones que son universales en las economías de mercado. Lo único que se requiere es la estimación de los parámetros que no se pueden determinar con la teoría económica. Estos parámetros se pueden calcular utilizando conjuntos de datos de una serie de zonas con características de comportamiento similares. También hemos aprendido que el software de este tipo de modelos tiene que ser fácil de utilizar y permitir al usuario verificar la racionalidad de los resultados.

El funcionamiento de los modelos de análisis del desarrollo regional/local utilizados para el pronóstico/simulación de políticas/programas en los Estados Unidos

En EU, los modelos económicos regionales y locales se utilizan para una serie de fines en el análisis de las políticas, incluyendo la evaluación de propuestas de desarrollo económico, proyectos de transporte, regulaciones ambientales y programas energéticos. Los modelos regionales también se utilizan con fines de planeación y programación, en particular en lo que se relaciona con necesidades de infraestructura, incluyendo nuevas carreteras, aeropuertos, plantas eléctricas, suministro de agua y una amplia gama de otros servicios públicos y privados.

Los profesionales utilizan una serie de métodos para los pronósticos y análisis de las políticas. Los modelos básicos son: insumo-producto, equilibrio general computable, econometría y nueva geografía económica. Cada método tiene sus fortalezas y debilidades. Como el modelo REMI integra todas estas metodologías, proporciona un sistema de trabajo para la modelación que supera algunas de las debilidades de utilizar solamente metodologías específicas de modelación.

Los modelos insumo-producto representan de qué manera opera la economía nacional, regional y local a través de la interacción de diversas partes de la economía. La interacción de los sectores industriales en la economía es central para el análisis insumo-producto. Por ejemplo, para construir un automóvil, el fabricante automotriz tiene que utilizar en la producción acero, llantas y otros insumos intermedios. La utilización de mano de obra y el capital también se consideran insumos de la producción, y la demanda final de consumo, gasto gubernamental y exportación también se representan en la contabilidad insumo-producto. Normalmente, los modelos insumo-producto incorporan con sumo detalle a un gran número de industrias y a veinte o treinta tipos de demanda final.

Los modelos insumo-producto tienen varias limitaciones importantes. Primera, parten del supuesto de que las relaciones económicas son fijas y que no varían en respuesta a cambios en los costos, a la competencia por los recursos o a las restricciones de la oferta. Segunda, son de naturaleza estática y no muestran los procesos dinámicos en el transcurso del tiempo. Tercera, a menudo son limitados en su descripción de la economía; por ejemplo, los modelos insumo-producto no toman en cuenta la interacción entre la población y los cambios económicos.

Los modelos de equilibrio general computable (CGE, por sus siglas en inglés) se suelen basar en fundamentos microeconómicos debidamente formulados: las firmas maximizan las ganancias y los hogares maximizan la utilidad. En contraste con los modelos insumo-producto, los de equilibrio general computable están estructurados de manera tal que los precios y salarios responden a las condiciones del mercado. Dado que la función de los indicadores de precio y costos en la

economía está plenamente especificada, los modelos CGE son particularmente idóneos para el análisis de políticas relacionadas con los cambios de impuestos y de costos de producción.

Los modelos de equilibrio general computable se utilizan para evaluar cambios en política comercial, regulaciones ambientales e impuestos. Por ejemplo, el modelo CGE puede utilizarse para mostrar los efectos económicos de un aumento del impuesto sobre la renta en el ámbito estatal. Los modelos CGE no se suelen aplicar a cuestiones de desarrollo económico, como el efecto de la ubicación de una nueva empresa, ya que generalmente carecen de detalles de la industria o de un atento seguimiento de las relaciones entre las industrias que sí contienen los modelos insumo-producto. Además, los modelos CGE, tal como se aplican comúnmente, a menudo carecen de una dimensión temporal explícita. Por lo tanto, el análisis de las políticas que utiliza dichos modelos asume la forma de un análisis comparativo estático que puede servir como un insumo útil para el proceso de diseño de la política, pero que tiene limitaciones prácticas importantes.

Los modelos econométricos utilizan rigurosamente métodos estadísticos. Dichos modelos tienen una base importante en la teoría económica y se fundan más en relaciones respaldadas empíricamente. Los modelos econométricos casi siempre tienen una importante dimensión temporal y dominan a otros métodos en los pronósticos a corto plazo trimestrales y mensuales. Los modelos econométricos se utilizan a veces con el propósito de analizar las políticas, pero no suelen ser idóneos para este fin ya que la estructura económica puede que no esté representada con suficiente detalle para los tipos de conmociones o *shocks* exógenos típicos en la aplicación de las políticas.

Los modelos de geografía económica se han desarrollado en su mayor parte en un contexto teórico en los diez últimos años. Como en el caso de los modelos CGE, este tipo de modelo también se basa en claros fundamentos microeconómicos pero normalmente es mucho más estilizado y no pretende representar una economía real. Los modelos de geografía económica son únicos en el sentido de que son capaces de explicar la formación endógena de ciudades y otras aglomeraciones económicas.¹

El modelo REMI reúne los métodos de los modelos de insumo-producto, de equilibrio general computable, econométrico y de geografía económica en un sistema congruente y exhaustivo. El modelo REMI capta las relaciones interindustriales incorporadas en los modelos insumo-producto. Si se deshabilitan todas las respuestas dinámicas en el REMI (como se puede hacer mediante la “especificación del modelo alternativo”), el modelo se prescribe como un modelo insumo-producto estático. El modelo REMI capta las tendencias a largo plazo del equilibrio general en los mercados de trabajo y de factores, como en los modelos de equilibrio general computable.

En el modelo REMI las respuestas dinámicas se calculan utilizando técnicas econométricas, normalmente utilizando cálculos basados en datos de panel correspondientes a varias regiones y estados, a lo largo de un número de años. El modelo también está basado en la nueva teoría de la geografía económica y explica claramente las economías de aglomeración en los mercados laboral, de insumos y de consumo.

Descripción de la estructura del REMI²

El modelo REMI de Percepción de las Políticas es un modelo estructural de pronóstico económico y análisis de las políticas. Como ya hemos mencionado, integra las metodologías de insumo-producto, de equilibrio general computable, econométrica y de geografía económica. El modelo es dinámico, con pronósticos y simulaciones generados anualmente en relación con las respuestas conductuales a salarios, precios y otros factores económicos.

El modelo REMI consiste en miles de ecuaciones simultáneas; sin embargo, su estructura básica es relativamente directa. El número exacto de ecuaciones varía dependiendo del alcance de la industria, la demografía, la demanda y otros detalles contenidos en el modelo. La estructura global del modelo se puede resumir en cinco bloques principales: 1) producto total y demanda, 2) demanda de mano de obra y de capital, 3) población y fuerza de trabajo, 4) salarios, precios y costos, y 5) participaciones del mercado. Los bloques y sus interacciones clave se muestran en las figuras 6.1 y 6.2.

El bloque de producto total y demanda consiste en producto total, demanda, consumo, inversión, gasto gubernamental, exportaciones e importaciones, además de la retroalimentación derivada del cambio en el producto total debido al cambio en la productividad de insumos intermedios. El bloque de demanda de trabajo y capital incluye la intensidad laboral y la productividad, así como la demanda de mano de obra y capital. Las ecuaciones del índice de participación de la fuerza de trabajo y de la migración están en el bloque de población y fuerza laboral. El bloque de salarios, precios y costos incluye las ecuaciones de los precios compuestos, de los determinantes de los costos de producción, del deflactor del precio de consumo, de los precios de la vivienda y del salario. La proporción de los mercados local, interregional y de exportación captados por cada región se incluye en el bloque de participaciones del mercado.

Se pueden construir modelos para una sola región, multirregionales o nacionales multirregionales. Se define ampliamente a la región como una zona subnacional y puede consistir en un estado, provincia, distrito o ciudad, o en cualquier combinación de zonas subnacionales. Dentro de una gran zona multinacional monetaria como la Unión Europea, se pueden construir modelos de una economía nacional aplicando el mismo marco económico empleado en los modelos regionales.

Figura 6.1 Vinculaciones del modelo REMI

Excluidas las relaciones de geografía económica

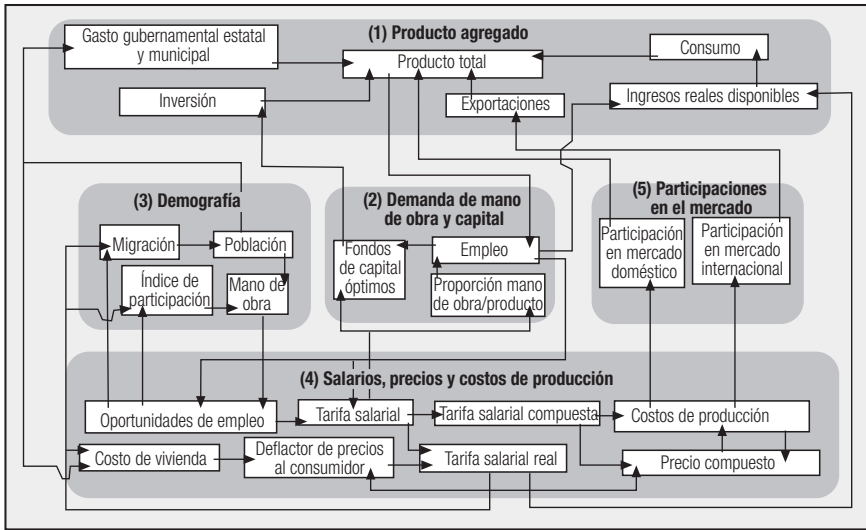
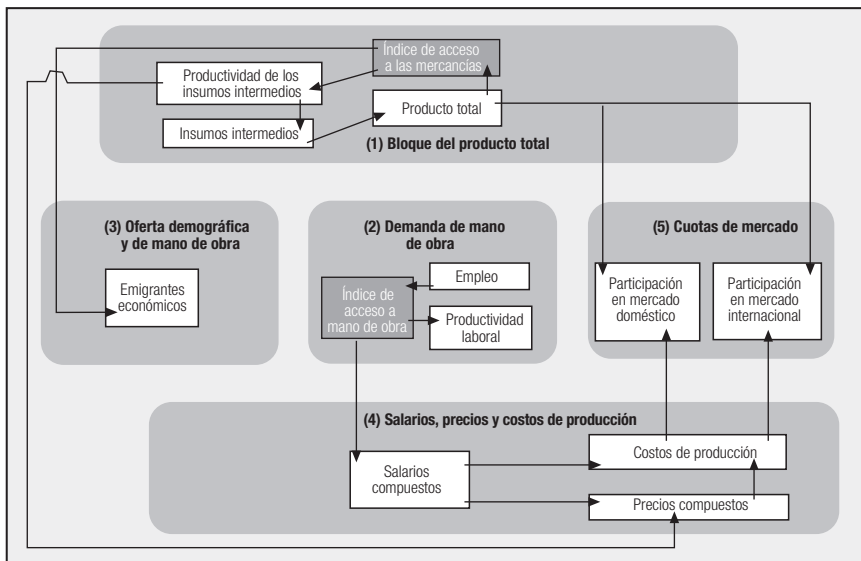


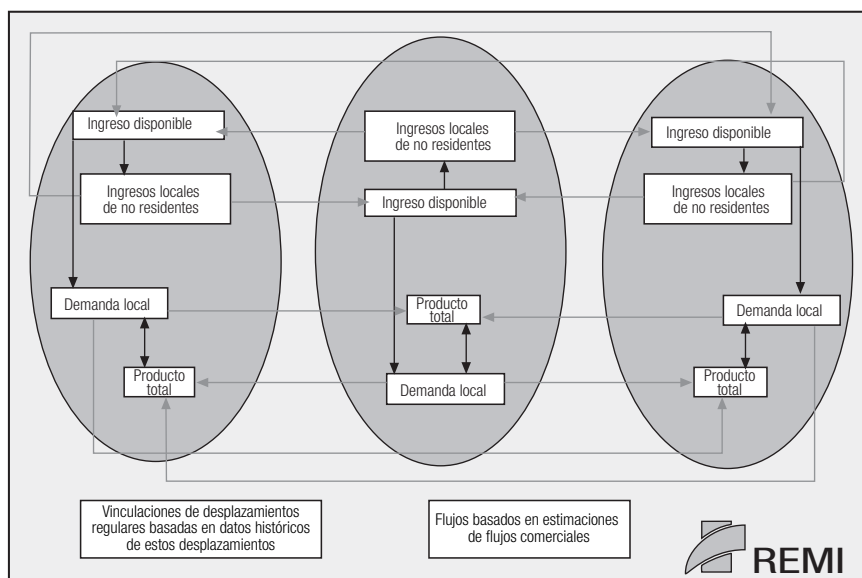
Figura 6.2 Vinculaciones de geografía económica



Los modelos de una sola región están en una región específica, denominada región hogar, pero el resto del país también está representado en el modelo. Sin embargo, como la región hogar es sólo una pequeña parte del total de la nación, los cambios en la región no tienen efectos endógenos sobre las variables del resto del país.

En los modelos multirregionales hay interacciones entre las regiones, como flujos comerciales y desplazamientos regulares de personas. Estas interacciones incluyen los flujos comerciales de cada región con cada una de las demás regiones. En la Figura 6.3 se ilustran estos flujos en un modelo de tres regiones. También hay vinculaciones multiregionales a través de los precios de entrega de una zona con cada una de las otras, así como del acceso a la mano de obra y las mercancías entre las zonas.

Figura 6.3 **Vinculaciones de los flujos comerciales y de desplazamientos regulares**



Los modelos nacionales multirregionales que abarcan toda una unión monetaria, tales como EU o la Unión Europea (UE), también pueden incluir una respuesta monetaria del banco central que restrinja los mercados de fuerza de trabajo. Los modelos que abarcan sólo una parte relativamente pequeña de una unión monetaria no están restringidos de manera endógena por variaciones en las tasas de cambio ni por respuestas monetarias.

Bloque 1. Producto total y demanda

Este bloque incluye los conceptos: producto total, demanda, consumo, inversión, gasto gubernamental, importación, acceso al producto y exportación. El producto de cada industria en la región hogar está determinado por la demanda industrial en todas las regiones del país, la participación en cada mercado de la región hogar y las exportaciones internacionales de cada región.

La demanda de cada industria está determinada por el monto del producto agregado, el consumo, la inversión y la demanda de capital de esa industria. El consumo depende del ingreso real disponible per cápita, los precios relativos, las elasticidades diferenciales del ingreso y la población.³ La productividad del insumo depende del acceso a los insumos porque cuanto mayor es el conjunto de insumos a elegir, más probable es que se forme el insumo con las características específicas requeridas para el empleo. En el proceso de ajuste de los valores de capital, la inversión cubre la diferencia entre los valores óptimos de capital y los reales para inversión residencial, no residencial y de equipo.⁴ Las variaciones del gasto gubernamental están determinadas por los cambios en la población.

Bloque 2. Demanda de mano de obra y de capital

El bloque de demanda de mano de obra y de capital incluye la determinación de la productividad laboral, la intensidad laboral y los valores de capital óptimos. La productividad de la mano de obra específica en cada industria depende de la disponibilidad de trabajadores con habilidades diferenciadas para las labores requeridas en cada industria. La oferta de mano de obra ocupacional y los costos de los desplazamientos laborales regulares determinan el acceso de las empresas a fuerza de trabajo especializada.

La intensidad laboral está determinada por el costo de la mano de obra con relación a los otros factores de insumo, el capital y el combustible. La demanda de capital está regida por la ecuación de valores óptimos del capital tanto para el capital no residencial como para equipo. Los valores óptimos de capital para cada industria dependen de los costos relativos de mano de obra y de capital, y del empleo ponderado por el capital utilizado en cada industria. El empleo en las industrias privadas está determinado por el valor agregado y el empleo por unidad de valor agregado en cada industria.

Bloque 3. Población y fuerza de trabajo

El bloque de población y de fuerza de trabajo incluye información demográfica detallada sobre la región. Los datos de la población se dan por edad, género y categoría étnica, con índices de natalidad y supervivencia para cada grupo. El tamaño y el índice de participación en la fuerza laboral de cada grupo determinan la oferta

de mano de obra. Estos índices de participación responden a cambios en el empleo en relación con la fuerza de trabajo potencial y a cambios en la tarifa salarial real después de deducir impuestos.⁵ La migración incluye la jubilación, los militares y la migración internacional y económica.⁶ La migración económica está determinada por la tarifa relativa de salario real con deducción de impuestos, las oportunidades relativas de trabajo y el acceso del consumidor a una oferta variada.

Bloque 4. Salarios, precios y costos

Este bloque incluye precios de entrega, costos de producción, costos de equipo, el deflactor del consumo, precios al consumidor, precios de la vivienda y la ecuación del salario. Los conceptos de geografía económica dan razón de los efectos que tiene en la productividad y el precio del acceso a mano de obra especializada, bienes y servicios.

Estos precios miden el precio del producto industrial, tomando en cuenta el acceso a los lugares de producción. Este acceso es importante debido a la especialización de la producción que se da dentro de cada industria y porque los costos de transporte y de transacción debidos a las distancias son importantes. Después, los precios compuestos para cada industria se calculan con base en los costos de producción de las regiones que suministran, la distancia efectiva a que están estas regiones y el índice de acceso a la variedad del producto en la industria en relación con el acceso de otros usuarios del producto.

El costo de producción para cada industria está determinado por el costo de la mano de obra, de capital, de combustible y de insumos intermedios. Los costos de la mano de obra reflejan un ajuste de la productividad para tomar en cuenta el acceso a mano de obra especializada, así como las tarifas salariales subyacentes. Los costos de capital incluyen los costos de estructuras no residenciales y de equipo, mientras que los costos de combustible incorporan la electricidad, el gas natural y combustibles residuales.

El deflactor del consumo convierte los precios industriales a precios de mercancías de consumo. Para los emigrantes potenciales, se calcula adicionalmente el precio al consumidor para incluir los precios de la vivienda. La variación respecto al nivel inicial de los precios de la vivienda depende de los cambios en el ingreso y de la densidad de población.

Los cambios salariales se deben a modificaciones en las condiciones de la demanda y la oferta de mano de obra y a variaciones en la tarifa salarial nacional. Los cambios en las oportunidades de empleo, en relación con los cambios en la fuerza de trabajo y la demanda ocupacional, determinan las tarifas salariales por industria.

Bloque 5. Participaciones en el mercado

Las ecuaciones de participación en el mercado miden la proporción de mercados locales y de exportación que capta cada industria. Estas participaciones dependen de los costos relativos de producción, de la elasticidad estimada del precio de la demanda y de la distancia efectiva entre la región hogar y cada una de las otras regiones. El cambio en la participación de una zona específica en cualquier región depende de las variaciones del precio de entrega y de la cantidad que produce en comparación con los mismos factores correspondientes a los competidores en ese mercado. La participación en los mercados local y externo después rige las exportaciones y las importaciones de la economía hogar.

Descripción de los datos

El modelo REMI tiene dos usos principales para los datos. Un uso es para los datos de series temporales, cuando se dispone de ellos, para estimar los coeficientes de las ecuaciones. El otro uso de los datos es para calibrar el modelo con el año que servirá como año histórico base.

Estimación de los coeficientes

Debido a la completitud de la estructura del modelo REMI, sólo hay un conjunto relativamente pequeño de coeficientes clave de las ecuaciones a ser estimados. Esto es posible gracias a que el modelo se apoya en la aplicación de la teoría económica de la corriente dominante, incluyendo la recientemente desarrollada geografía económica. Los supuestos clave, subyacentes en el modelo, son que las firmas buscan maximizar las ganancias, los hogares buscan maximizar su bienestar, y que los trabajadores, así como los bienes y servicios, no son homogéneos ni siquiera cuando están en la misma ocupación o sector de la economía. La estructura de ecuaciones del modelo se deriva de estos supuestos básicos. En nuestros más de veinte años de experiencia, se ha vuelto evidente que, debido a la calidad de los datos regionales, la complejidad de la economía y la información que las empresas y los hogares consideran, mientras mayor es el conjunto de datos utilizado mejor es la calidad de las estimaciones. Esto es cierto incluso después de que la estructura del modelo ha sido diseñada de tal manera que el número de coeficientes y los requerimientos de datos han sido reducidos al mínimo.

Para la cuantificación en este nivel básico, la similitud del comportamiento para los actores en las diferentes regiones es más importante para una estimación robusta que la importancia de las pequeñas diferencias de una región a otra. En los Estados Unidos esto nos ha llevado a desarrollar una base de datos al nivel de 53 industrias por más de 25 años para 3083 condados. Esta base de

datos incluye obviamente los condados que abarcan un espectro que va desde pequeños condados rurales con sólo unos miles de personas hasta condados urbanos densamente poblados. En el proceso de estimación, filtramos los datos para asegurar la precisión para la industria en cada condado que está incluido en la estimación específica de la industria.

En Europa, los conjuntos de datos de las series temporales son más escasos y a menudo de un ámbito geográfico más alto pero nos proporcionan un conjunto de datos que hace posible cualquier ajuste necesario de nuestras estimaciones de coeficientes para los Estados Unidos debido a la condiciones específicas de la Unión Europea.

En el resto de esta sección expondremos todas las estimaciones necesarias de los coeficientes de las ecuaciones en grupos lógicos.

Iniciamos con dos coeficientes clave utilizados en las ecuaciones de participación de mercado.

β_i = El parámetro de decaimiento por la distancia (medida como el tiempo de transporte) en un modelo de gravitación para la industria i .

σ_i = La elasticidad del precio para la industria i .

El coeficiente β_i determina cómo declina la participación de los territorios en la demanda de mercado en cada uno de los otros territorios en la medida en que el tiempo de transporte a cada uno de los otros territorios aumenta. El valor de σ_i muestra cómo cambiará la participación del territorio hogar en cada mercado (incluyendo su propio mercado) en respuesta a un cambio en el costo de entrega del territorio hogar a cada uno de los otros territorios. Esto se relaciona con el cambio promedio del costo de entrega en cada territorio con base en sus compras a todos los otros territorios de los cuales está comprando. El cambio en el costo de entrega desde el territorio hogar puede ser debido a cambios en sus costos de producción o cambios en el costo de entrega.

Con el propósito de estimar dinámicamente el β_i (el parámetro de decaimiento por tiempo transporte), necesitamos tener estimaciones por series temporales de lo siguiente:

1. Una aproximación del cambio en el producto total de la industria para los bienes y servicios producidos domésticamente en cada territorio, cada año por cada industria.
2. Una aproximación del cambio en la demanda total de bienes y servicios producidos domésticamente en cada territorio para todas las industrias y todos los años. Esto se descompone en:

a) Demanda de los consumidores.

- b) Demanda de inversión (cambio del capital bruto).
- c) Demanda gubernamental.
- d) Demanda intermedia.

Podemos aproximar las variaciones en cada uno de estos conceptos si tenemos los cambios en el empleo por industria, una matriz insumo-producto nacional (o de la Unión Europea), y una matriz de tiempos de transporte para cada territorio a cada uno de los otros territorios. Nosotros hemos obtenido dicha matriz del profesor Wegener, comisionada por la UE para todas las regiones NUTS II y NUTS III* en las zonas de la UE.⁷ Esto se logra mediante lo siguiente:

1. Una aproximación de las variaciones en el producto total destinado al uso doméstico en cada territorio por año y por industria: utilice los productos nacionales por industria de la matriz insumo-producto nacional y los datos del empleo nacional (o utilice, si están disponibles, series de tiempo del producto nacional por industria y del empleo por industria), luego aplique la razón apropiada a las series del empleo local.
2. Una variación en la demanda total por industria, por territorio y por año.
 - a) Cambio en la demanda de consumo: con base en las variaciones en el salario facturado (o el ingreso disponible si existe la cifra), convertido a demanda industrial utilizando el vector de consumo en la matriz insumo-producto nacional o de la UE.
 - b) Cambio en la demanda de inversión: con base en el salario facturado en la construcción (o variaciones en la tasa de cambio en el Inventario del Capital Bruto) convertido a industria utilizando la variación en el inventario del capital bruto en la matriz insumo-producto.
 - c) Demanda gubernamental: variaciones en el salario total (o variaciones en la población) convertida a demanda gubernamental de la economía privada a demanda de la industria utilizando la matriz insumo-producto apropiada.
 - d) Cambios en la demanda intermedia del producto de cada industria con base en los productos agregados estimados arriba en (1) y la matriz insumo-producto nacional (o de la UE).

Después de integrados estos datos, entonces el β_i (parámetro de decaimiento por tiempo transporte) se puede encontrar utilizando un algoritmo de búsqueda entre valores de β_i para encontrar los valores β_i apropiados. Esto minimiza el índice de errores para predecir las variaciones del producto total en cada territorio, con base en las variaciones de la demanda en cada uno de los territorios para los

* NUTS representa las siglas de Nomenclature of Territorial Units for Statistics [Nomenclatura de Unidades Territoriales para Estadísticas] en la Unión Europea: http://ec.europa.eu/comm/eurostat/ramon/nuts/home_regions_en.html

bienes o servicios de tal industria. Por ejemplo, el producto de las tiendas de alimentos debería estar estrechamente relacionado (i.e. en gran proporción) con las variaciones en la demanda en la zona hogar y áreas contiguas, y no muy relacionado (i.e. en baja proporción) con las variaciones en la demanda de las zonas distantes. Sin embargo, en el territorio hogar el producto de la industria automotriz debería estar mucho más relacionado con las variaciones en la demanda en todos los territorios que con las variaciones en la demanda de la zona hogar (i.e. muy bajo decaimiento) debido al tiempo de transporte. Esto producirá una participación más o menos igual en todos los mercados.

Es posible una manera alternativa de estimar los valores de los β_i , si existen una o más matrices insumo-producto subnacionales para cualquier año. Este método proporciona otra estimación cuantitativa para ajustar los valores de los β_i con el propósito de explicar mejor los flujos entrantes y salientes de la región local con las otras regiones con base en un año de datos. En los EU, examinamos las implicaciones de cada uno de los valores de los β_i para el tiempo promedio de transporte para el bien o servicio en cuestión en relación con su costo. También observamos la proporción de la demanda de la zona local servida por el producto local sobre un rango de zonas. Comparamos las proporciones con las de industrias similares. Estimamos el tiempo de transporte de los clientes típicos para el bien o servicio específico en cuestión. Después de hacer este análisis, usamos toda la información para modificar algunos de los valores de los β_i .

La estimación de la σ_i (respuesta de la elasticidad del precio) utiliza los mismos datos de series de tiempo dispuestas en el enfoque de series temporales para las estimaciones de los β_i señalados antes. En este caso, utilizamos un algoritmo para buscar entre los valores de σ_i para encontrar qué valor de σ_i mejoraría el ajuste entre las variaciones en el producto y las variaciones en la demanda con base en los cambios en las participaciones que el coeficiente predeciría en los mercados subsecuentes a la variación relativa en los costos de producción (aproximado mediante las variaciones en la tarifa salarial) en todas las zonas. En otras palabras, la elasticidad de la demanda σ_i (el cambio porcentual en la cantidad demandada dado el cambio porcentual en los precios relativos de entrega) se determinaría de forma econométrica.

En los Estados Unidos, hemos reunido las empresas similares en categorías de manera que la elasticidad estimada es idéntica para todas las industrias en la categoría general. También analizamos las elasticidades para que sean razonables. Para algunas de las industrias, cambiamos el criterio del filtro para depurar los territorios que no tienen una representación sustancial de la industria en cuestión o que tienen series de tiempo erráticas, posiblemente debido a errores en el reporte de los datos o su clasificación. Utilizando estos métodos, obtenemos estimaciones estadísticamente significativas para todas las industrias que cumplen nuestra prueba de razonabilidad profesional.

Observando los datos del empleo y de los salarios disponibles y las matrices insumo-producto existentes para los países de la Unión Europea, parece que no tendremos problemas en construir modelos para esas regiones.

Nosotros satisfacemos típicamente los requerimientos de datos para los modelos de la UE de las siguientes maneras:

1. A nuestros socios en la empresa conjunta se les proporciona una lista de nuestros requerimientos y se les solicita proveer todos los datos que puedan proporcionar.
2. Acudimos a la base de datos New Cronos del Eurostat para los datos que no nos proporcionan nuestros socios.
3. Extraemos datos de internet y hacemos nuestras propias solicitudes a cualquier otra fuente disponible. No obstante, incluso con estas tres fuentes, muchas veces nos faltan algunos puntos y nos vemos en la necesidad de cubrir estos vacíos utilizando procedimientos de estimación.
4. En los casos que están disponibles datos de años previos, ajustar una curva a los datos disponibles y extenderla al año de interés estima los datos para el año histórico de interés. Cuando los datos suministrados son para un ámbito regional superior al requerido, los datos disponibles se distribuyen utilizando variables correlacionadas que reflejen las variaciones regionales. Se hace una distribución similar cuando los datos disponibles son un agregado de las industrias requeridas. En dichos procedimientos de estimación, normalmente tenemos totales de control en el ámbito geográfico nacional, en el ámbito subnacional y en un cierto agregado de industrias.

El Modelo REMI para la Percepción de Políticas Públicas está proyectado para utilizar datos demográficos detallados, a los que tenemos fácil acceso en los Estados Unidos. En muchos países europeos, no existen datos al nivel de detalle que necesitamos. Por lo tanto, hay casos en los que se requieren métodos estadísticos para llenar las partes de los datos faltantes, y tenemos muchos medios a nuestra disposición para hacerlo.

Generalmente, la población por cohortes de edad individuales es de fácil acceso, al menos en el ámbito nacional. Si tenemos esta información, entonces utilizaremos los datos para distribuirlos en el ámbito regional, si es necesario, suponiendo que en el ámbito regional cada grupo de edad dentro de una cohorte de cinco años estará en la misma proporción recíproca que la que tienen en el ámbito nacional.

Si no obtenemos el dato de la fuerza de trabajo local, o no tenemos un pronóstico del índice de participación, utilizaremos los índices de participación predichos por Eurostat. Como Eurostat no tiene el mismo número de grupos de

clase o cohortes que queremos, tenemos que distribuir estos datos. Lo que hemos hecho en el pasado es calcular los índices de participación a partir de la población y de la fuerza de trabajo de Eurostat. Luego utilizamos los mismos índices de participación para cada uno de los grupos o cohortes de 16 a 19 años de edad y asimismo para cada cohorte individual de edad de más de 75 años. Después utilizaremos las agrupaciones de género y edad de 5 años para las cohortes restantes.

En el futuro, planeamos usar la población de Eurostat y distribuirla utilizando los datos de la población que ya tenemos. Después usaremos la fuerza laboral de Eurostat y la distribuiremos, utilizando datos reales de la fuerza laboral de un país cercano o las proporciones de la fuerza laboral en los EU. para repartir los grupos de clase o cohortes con rangos de 5 años en grupos individuales de clase por edad. Entonces podremos calcular los índices de participación utilizando estos valores. A continuación, utilizaremos los índices de participación calculados por Eurostat con los datos de la población suministrados por el cliente para calcular la fuerza laboral del pronóstico. Eurostat sólo proporciona predicciones de la población y de la fuerza laboral para periodos cada cinco años, de manera que utilizamos un método lineal para llenar los índices de participación para los años intermedios.

Para los datos de la migración, utilizaremos datos nacionales para distribuir los grupos de clase en el ámbito regional, o podemos utilizar los datos de los EU u otro país distribuidos tanto en el ámbito nacional como regional si es necesario.

Si carecemos de la predicción del índice de natalidad, suponemos que los índices de natalidad permanecen constantes en el periodo de tiempo pronosticado.

Para el empleo a través de datos de la actividad ocupacional, llenamos los porcentajes para una industria faltante mediante la utilización de los valores de la industria más similar. Cuando no tenemos el desglose de las industrias de manufactura, utilizamos los mismos valores para todas las industrias de manufactura que se nos proporcionan para la industria de manufactura en su conjunto.

Una vez utilizados los datos disponibles para estimar los β_i y las σ_i en el país en cuestión, examinaremos estas estimaciones a la luz de las estimaciones para Estados Unidos y otras estimaciones de la UE para verificar la consistencia y la razonabilidad. Nuestros criterios para examinar las estimaciones son los siguientes: el valor estadístico de las estimaciones (*v. gr.* sus valores de prueba *t*-student), la similitud de las estimaciones con las de EU y otros países de la UE, la consistencia de los resultados de una industria a otra, y la opinión de nuestros contactos en el país en cuestión. Utilizamos nuestro juicio profesional para determinar la aplicación directa de las nuevas estimaciones, la utilización de otros grupos de combinaciones, o para encontrar otras formas de llegar a las

mejores estimaciones posibles. Si no es posible estimación alguna para un país en particular debido a la carencia de datos, utilizaremos otras estimaciones de los β_i y las σ_i en la UE combinadas con las de Estados Unidos para el país en cuestión. Dado el dominio de muchas industrias por firmas internacionales, existe una gran probabilidad de que las señales de las elasticidades del consumo de la UE sean bastante consistentes con sus señales en Estados Unidos, por lo que puede ser razonable utilizar las estimaciones de Estados Unidos en algunos casos. Gracias a nuestras técnicas de estimación, se puede hacer una estimación de serie temporal con tan sólo dos o tres años de datos. Por consiguiente, siempre tendremos una forma de probar nuestras estimaciones en todos los países que estudiamos para asegurar que nuestros resultados son consistentes con otros países.

La ecuación de la productividad de la mano de obra utiliza las σ_i (estimaciones de la heterogeneidad) del costo de la mano de obra, las cuales luego son utilizadas en el índice de acceso a la mano de obra. Estas estimaciones se hicieron con base en el monto y el costo del flujo cruzado entre comunidades en 1300 Zonas de Análisis de Tráfico (ZAT's) alrededor de Chicago. Los datos incluyeron la ocupación y la industria, así como el lugar de residencia y el lugar de trabajo de todos los incluidos en la muestra. También obtuvimos estimaciones del tiempo de transporte y los costos de los flujos entre comunidades de cada ZAT a cada una de las otras ZAT. Las estimaciones son bastante robustas y como representan las decisiones de contratación de la industria, se espera que sean similares en cualquier economía en la que haya beneficios en el acoplamiento de los trabajadores a su trabajo para beneficio tanto de las empresas como de los trabajadores.⁸

La velocidad de respuesta de la migración económica es una ecuación clave para estimar los datos europeos. Con el propósito de estimar esta respuesta, requerimos series de tiempo sobre las tarifas salariales reales, preferiblemente utilizando un deflactor que incluya los precios de la vivienda, la razón relativa del empleo con la fuerza laboral, y el número neto de emigrantes internos anuales de región a región. Estas series, con la posible excepción del índice de precios que utiliza los precios de la vivienda, son necesarias para probar alguna diferencia esperada en la velocidad de ajuste de Estados Unidos y la UE. En caso de que se carezca de datos para las series de tiempo, podríamos establecer si la velocidad de la reacción de ajuste se debería modificar respecto a la de Estados Unidos u otros países de Europa mediante la observación de los datos de flujos migratorios que existan.

La ecuación de la fuerza laboral tiene dos conjuntos de coeficientes. El primer conjunto está relacionado con las respuestas de participación de la fuerza laboral. Esto refleja cambios en el empleo actual de las regiones en relación tanto con una fuerza laboral sintética utilizando los índices de participación nacional por grupos de clase, como con la población actual por grupos de clase en edad/género. El otro conjunto se relaciona con la tarifa relativa de salario real. Si carece de variaciones,

el índice de participación seguirá con lo predicho en la nación para el punto de partida. Los coeficientes se dan por grupos de clase con rango de 20 años para los dos géneros; estos coeficientes se podrían calibrar colectivamente para cada país si hay índices de participación masculina y femenina de varios años para las regiones modeladas. También podría haber estudios europeos de variaciones en el índice de participación causadas por variaciones en la tarifa del salario real y el índice de desempleo. Si existen estos estudios, se podrían incorporar.

Cuando los índices de desempleo son altos, la señales generales de los bajos índices de participación capturan (y el tiempo de ajuste a estos cambios refleja) los efectos de largo plazo causados por un alto y continuo desempleo en algunas regiones. En el agregado, quisiéramos actualizar esta ecuación basada en las variaciones regionales en Europa (o en el país en cuestión) del índice de participación con base en las variaciones en las condiciones de desempleo y la tarifa salarial real. Si es posible, también podemos probar los efectos de las variaciones en el salario real en relación con los subsidios disponibles para los que no tienen trabajo.

La ecuación del precio de la vivienda (o de la tierra) se estimará con datos provenientes de la zona en cuestión. El cambio en el precio depende de los cambios en el ingreso real disponible relativo y de la densidad de población relativa. La serie más difícil de encontrar es la estimación del precio de la vivienda, que se tiene que estimar utilizando la serie compilada por la industria inmobiliaria.

Finalmente, la ecuación del salario depende del promedio móvil relativo del empleo dividido entre el promedio móvil de la fuerza laboral y la demanda ocupacional actual ponderada por industria entre su promedio móvil. Los datos de esta ecuación requieren de los datos de las industrias y del empleo mencionados en la primera parte de esta sección, además de una matriz de razones del personal por industria/ocupación a escala nacional. Estimaciones iniciales presentan una respuesta salarial algo menor en la UE que en Estados Unidos. La respuesta salarial requeriría, de ser posible, ser estimada de nuevo para los países en cuestión.

Resumen para la estimación de coeficientes

A partir de la evidencia disponible, parece que existen los datos adecuados para estimar los coeficientes necesarios para la construcción de modelos para Europa, Norteamérica y la mayoría de los otros países de mercado. Cuando no se pueden estimar coeficientes en particular para un determinado país, tendremos evidencia cuantitativa de otros países similares que se podría utilizar para modificar los coeficientes de Estados Unidos como se requiere. Consideramos necesario para trabajos ulteriores, encontrar y extraer datos ya existentes pero que no son fáciles de obtener.

Calibración del modelo

Aunque REMI ha reducido el número de coeficientes que se necesita estimar a un nivel mínimo, la calibración del año histórico anterior designado continúa siendo una tarea adicional. Implica utilizar toda la información que ya hemos mencionado, pero sólo para un año. También implica hacer un pronóstico nacional del país o unión monetaria en la que se ubica la región. Podemos hacer este pronóstico nosotros mismos, pero, si es necesario, hemos de estar dispuestos a alinearlo con un pronóstico oficial.

Para hacer nuestro pronóstico nacional, primero desarrollamos una proyección de la fuerza laboral nacional mediante la aplicación de los índices de participación a los grupos de clase que sean consistentes con los índices de participación proyectados por Eurostat. La combinación de nuestro crecimiento de la productividad por industria proyectado con el tamaño de nuestra fuerza laboral impone un límite de producto agregado para mantener una razón fija del empleo/la fuerza laboral para el punto de partida del pronóstico nacional. Se podría implementar una tendencia a la alza en la proyección del empleo/fuerza laboral si fuera deseable suponer que los índices de desempleo actuales exceden sus probables tasas a largo plazo.

Hemos producido modelos de países europeos con 24, 26 y 30 industrias. En cada caso, acudimos al tamaño de región más pequeño posible (usualmente NUTS II o NUTS III) para preparar una base de datos que abarcara a todo el país. Esto nos permite estimar convenientemente los flujos comerciales en ese ámbito geográfico. Luego agregamos estos flujos para un número limitado de ecuaciones y calculamos el valor de un concepto, como la distancia efectiva (tiempo de transporte), de tal manera que los flujos internos y los de una región grande a otra serán consistentes con los flujos entre regiones más pequeñas.

La información demográfica necesaria para el pronóstico incluye detalles que requerimos solamente para la región que se está modelando. Esto se debe a nuestra necesidad de tener un modelo demográfico estructurado con todos los procesos demográficos estándar.

Costo de la construcción/uso del modelo

Como se desprende de la exposición anterior, el modelo REMI® de Percepción de Políticas Públicas es diferente y está diseñado a la medida para cada zona para la que construimos un modelo. Actualmente, la gama de nuestros modelos europeos va de 24 a 30 sectores industriales. Estos incluyen un modelo de la región de los Grampian en el Reino Unido, un modelo de dos zonas en Francia y el resto de la Unión Europea, un modelo de tres zonas en los Países Bajos, un modelo de dos zonas en el Rin-Westfalia, y un modelo de zona única en la Toscana. Los

tres últimos modelos involucraron a ECORYS, RWI (Rhine-Westphalia Institute for Economic Research, Essen) y a IRPET (Istituto Regionale Per La Programmazione Economica Toscana), respectivamente. REMI ha sido elegido recientemente por la Comisión Europea para desarrollar nuestra metodología de modelo para valorar el impacto regional de los fondos para la estructura; construiremos un modelo para Italia Meridional y un modelo de cuatro regiones contiguas en España como ejemplos para la Comisión Europea.

El modelo REMI de Percepción de las Políticas Públicas es un producto con licencia adecuado a las necesidades de la región o de las regiones a las que está dirigido. El precio es de €46 000 euros para un modelo de una sola zona de cualquier tamaño de región. El precio de un modelo de dos zonas es de €53 000 euros. Para cada región adicional, hasta diez regiones, es de €7 000 euros y llega a un precio de €112 000 euros para un modelo de 10 zonas. Por cada una de las 10 zonas siguientes el incremento es de €3 500 euros. Después del primer año, un costo de mantenimiento del 30% brinda un nuevo modelo cada año y la ayuda telefónica y por correo electrónico ilimitada para el uso del modelo y la interpretación de los resultados.

La utilización de las evaluaciones de programa o los efectos directos proyectados ex ante como insumos para el modelo REMI

El modelo REMI® de Percepción de Políticas Públicas es un instrumento de pronóstico macroeconómico y de análisis de las políticas. Incluye un pronóstico del punto de partida y proporciona variables de políticas públicas tales como el empleo, la productividad, los impuestos y los costos de producción que pueden ser modificadas por el usuario. Los analistas pueden obtener los valores de entrada para las variables del programa de evaluación de las políticas a partir de las salidas de evaluaciones limitadas del programa. Los analistas pueden obtener información similar de propuestas de intervenciones directas para proyectos que están en consideración.

Por ejemplo, los resultados directos del tradicional análisis costo-beneficio de inversiones para el transporte proporcionan los valores de entrada de las variables de las políticas necesarios para los estudios del transporte. Estas variables directas incluyen factores tales como gastos de construcción, gastos de operación y ahorros en tiempos de transporte para las empresas y los consumidores. Estas variables se pueden usar como insumos en el modelo REMI para mostrar los efectos macroeconómicos totales de las mejoras en la infraestructura del transporte.

El modelo REMI también se ha vinculado con modelos microeconómicos especializados. El modelo Energy 2002,⁹ que proporciona un alto nivel de información detallada en relación con la energía, permite a los usuarios simular

los efectos de diferentes tipos de plantas generadoras de energía eléctrica y de otras políticas energéticas. Las salidas producidas por Energy 2002, incluyendo el gasto en la construcción de servicios públicos así como las variaciones en el precio de los combustibles para el consumidor, la industria y el comercio, después se utilizan como información de entrada en el modelo REMI. A su vez, el modelo REMI proporciona las variaciones pronosticadas en los resultados que requiere el modelo Energy 2020.

Panorama de las áreas de análisis de las políticas públicas

El modelo REMI es un modelo exhaustivo de pronóstico y análisis de políticas. Incluye miles de variables económicas y demográficas de las políticas, así como una descripción completa de la economía regional. Por lo tanto, los analistas pueden usar el modelo para evaluar un amplio rango de opciones de políticas. Éstas incluyen programas e incentivos para el desarrollo, inversiones para el transporte, regulaciones ambientales y energéticas, y otras políticas que tienen efectos sobre la economía. Esta sección describe aplicaciones en muchas áreas importantes, con unos cuantos ejemplos ilustrativos que se basan en los miles de análisis de políticas públicas que se han llevado a cabo utilizando el modelo REMI.

Desarrollo económico

Una amplia variedad de programas de transferencias gubernamentales, inversiones en infraestructura y programas de incentivos empresariales se diseñan con el propósito de impulsar el desarrollo económico en zonas locales y regionales. El desarrollo económico como objetivo gubernamental, apunta a menudo a la atracción o retención de empresas en específico. Además, las mejoras de infraestructura (como por ejemplo, la ampliación de un aeropuerto) tienen a menudo un importante componente de desarrollo económico.

El modelo REMI se ha utilizado para una amplia gama de proyectos de desarrollo económico; éstos incluyen aspectos que abarcan desde los efectos de la industria de carreras y cría de caballos en Minnesota,¹⁰ el impacto de un centro de convenciones y hotel en Kansas City,¹¹ hasta el impacto económico de propuestas de juegos de azar en casinos en varios estados.¹² Los analistas han evaluado los efectos económicos de la ubicación de las armadoras de automóviles en Kentucky,¹³ Michigan,¹⁴ Illinois, Carolina del Norte y Texas. Otro tema es el de los efectos económicos de los estadios deportivos; los estudios incluyen un complejo futbolístico en Hartford, Connecticut y un nuevo parque de beisbol en Boston, Massachussets.¹⁵

Los diseñadores de políticas evalúan a menudo los efectos económicos de las pérdidas potenciales de industrias clave o de empleadores. Por ejemplo, los

usuarios de REMI han evaluado los efectos económicos regionales del cierre de bases militares,¹⁶ el decremento en las ventas de tabaco,¹⁷ las ventas perdidas de carbón,¹⁸ y el cierre de plantas industriales.¹⁹ Investigadores han evaluado los efectos económicos de los controles del uso de suelo y el crecimiento en California. Los usuarios de REMI también han evaluado las pérdidas económicas que podrían ocurrir por desastres naturales reales o potenciales, tales como inundaciones y huracanes.²⁰

Para evaluar una nueva empresa, el analista necesita normalmente considerar la fase de construcción, los requerimientos de la operación y los incentivos fiscales o de gasto gubernamentales que podrían haber sido necesarios para atraer a la empresa. La construcción de una instalación se da en un periodo de tiempo relativamente breve y usualmente genera muchos empleos temporales. Los analistas pueden ingresar la construcción en el modelo REMI utilizando, ya sea una variable de política general de construcción, ya sea variables de las políticas relacionadas con tipos específicos de construcción como nuevos edificios industriales, nuevos edificios de oficinas o nuevos edificios comerciales.

Los usuarios ingresan los efectos directos de la fase de operaciones de una nueva compañía utilizando variables de empleo o de producto de la política. Cuando usa la variable de la política de empleo, el analista ingresa el número total de trabajadores que están empleados directamente por la nueva empresa. Las variables de política de empleo suponen un cierto nivel de productividad laboral, así el modelo agrega el producto directo asociado con estos empleados. Asimismo, cuando el usuario ingresa valores de producto, el modelo calcula el empleo directo adicional que se necesita para generar este producto. El usuario ingresa los valores de empleo o producto directos antes de poner a funcionar el modelo para mostrar los efectos económicos totales de la nueva empresa.

Como parte de una política de desarrollo económico dirigida, las agencias gubernamentales ofrecen a menudo inversiones en infraestructura o incentivos fiscales específicos, o ambos, como una manera de atraer empresas. Por ejemplo, el estado puede proporcionar una rampa de salida de una carretera interestatal como parte de un paquete para atraer una instalación de manufactura. Lo más común es que los estados y las ciudades ofrezcan beneficios fiscales sobre ingresos corporativos, bienes, y de otros tipos dirigidos a atraer la construcción de una instalación en el estado.

Para captar adecuadamente los efectos económicos de la atracción de empresas, el analista requiere considerar explícitamente los efectos de las inversiones del gobierno o los incentivos fiscales. Para una inversión gubernamental en infraestructura, el analista requiere agregar el gasto en infraestructura y reducir el gasto gubernamental o el incremento en los impuestos en otras partes para pagar la inversión en infraestructura. Para un incentivo fiscal, el analista requiere

balancear el presupuesto gubernamental a través del incremento impositivo o la reducción del gasto gubernamental en otras partes con el propósito de compensar el recorte fiscal.

En algunos casos, los funcionarios públicos de desarrollo económico pueden aseverar que un incentivo fiscal para la empresa en forma de exención fiscal no implica necesariamente que los impuestos tengan que subir o que el gasto gubernamental deba reducirse en otras partes. Estos funcionarios argumentan que, como las exenciones fiscales no implican transferencias gubernamentales, no son necesarios otros cambios en el gravamen fiscal o en el gasto del gobierno. Para la mayoría de las exenciones fiscales, este argumento es falaz. En general, el suministro de servicios gubernamentales está estrechamente ligado a la actividad económica. Incluso las empresas que tienen bajas necesidades de servicios gubernamentales a menudo emplean a trabajadores que, como residentes, requieren un alto nivel de servicios públicos de educación, incendios, seguridad y otros. Así, un incremento en la actividad económica requiere un incremento en la recaudación para poder respaldar la necesidad de un nivel superior de servicios gubernamentales. Sin información ulterior, el analista debería suponer que los ingresos adicionales generados por la empresa son necesarios, al menos en parte, para pagar los requerimientos adicionales de servicios gubernamentales.

Los usuarios del modelo REMI pueden implementar cambios en la infraestructura gubernamental utilizando variables de política para inversiones detalladas en carreteras, en instalaciones marítimas y aéreas, y en otras categorías detalladas del gasto. Los usuarios pueden representar los cambios impositivos por el gobierno mediante variables de política para impuestos a los ingresos personales, impuestos a la propiedad, impuestos a ventas (para 13 mercancías de consumo), y una variedad de impuestos empresariales.

Algunas iniciativas de desarrollo económico poseen aspectos especiales que se pueden evaluar usando el modelo REMI. Un gran evento periódico o único, como un festival o un evento deportivo, a menudo se lleva a cabo como un medio de crear trabajos y actividad económica en una ciudad o región. La Universidad de Connecticut evaluó los efectos económicos de la propuesta de los Juegos Olímpicos Especiales de 1995 en New Haven, Connecticut.²¹ Dado que mucha de la construcción para preparar el evento y del empleo durante el evento son temporales, los mayores efectos económicos son temporales. Los efectos de largo plazo son resultado de las instalaciones adicionales que quedan disponibles para la región, así como los posibles efectos de la ubicación empresarial causados por la publicidad generada por el evento.

Las universidades y otras instituciones se suelen considerar como catalizadores para el desarrollo económico de una región, no solamente debido al empleo y el gasto inmediatos de la universidad, sino también a través del incremento de la

productividad de la fuerza laboral y su actuación como incubadora de nuevas tecnologías y empresas. Nexus Associates evaluó los efectos de la Escuela de Medicina Veterinaria de la Universidad Tufos en las economías de Boston y Massachusetts; en particular, el estudio indagó los efectos económicos de las nuevas divisiones de negocios creadas a partir de la investigación en la Universidad.²²

Algunas actividades económicas a las que se podría considerar no atractivas para una región se podrían valorar debido a sus beneficios para el desarrollo económico. La explotación de minas, las plantas generadoras de energía eléctrica, las fábricas de pulpa de papel y otras actividades dañinas para el ambiente, no obstante, se pueden valorar por su potencial de creación de empleos. Las prisiones, que suelen ser vistas como externalidades negativas debido a preocupaciones de seguridad, son buscadas intensamente como un motor del desarrollo económico por las regiones rurales deprimidas económicamente. Por ejemplo, el Departamento de Análisis Económico del gobernador de Kentucky evaluó los efectos económicos de la ubicación de una cárcel en diversas localidades de las regiones rurales montañosas del este de Kentucky. Las comunidades de esta región pretendieron activamente conseguir que se estableciera la prisión en sus zonas respectivas por el empleo estable que proporciona este tipo de establecimiento, empleo que de otra manera no existiría en el Kentucky de los Apalaches.²³

Transporte

Los analistas utilizan el modelo REMI para evaluar los efectos económicos totales de los proyectos de transporte. Este análisis se usa a menudo como medio para ir más allá de las estimaciones tradicionales de costo y beneficio directos y mostrar las repercusiones más generales en la región. Los analistas han utilizado el REMI para cientos de estudios de transporte que involucran autopistas, líneas ferroviarias, puertos y aeropuertos.

Wilbur Smith Associates utilizó el modelo REMI para llevar a cabo una evaluación *ex post* del sistema de Desarrollo de los Apalaches para Autopistas. Este sistema se construyó durante un periodo de más de treinta años con el propósito de incrementar el acceso a las comunidades remotas en las montañas de los Apalaches en Estados Unidos. Los analistas utilizaron las salidas de información generadas por modelos de transporte detallados como datos de entrada para el modelo REMI. Los cambios directos, implementados como variables REMI de la política, incluían variaciones en el costo competitivo, ventas generadas a lo largo de las carreteras y los gastos de viajes y el turismo. Los efectos económicos totales de estos cambios directos incluyeron estimaciones año por año de los incrementos en el producto bruto regional, el empleo y los salarios.²⁴

Un estudio sobre un enlace por ferrocarril de alta velocidad entre San Diego y San Francisco, California, empleó el modelo REMI para mostrar la factibilidad

del proyecto con base en mediciones beneficio/costo y valor presente neto.²⁵ Los investigadores consideraron diversos efectos directos de los cambios en la política y los modelaron utilizando las variables REMI adecuadas de la política. Para representar los efectos directos del gasto, se utilizaron variables detalladas de la política de construcción específica del ferrocarril y de su operación. Ya que la línea ferroviaria sería utilizada en lugar de otras formas de transporte, reduciendo así las congestiones, los analistas redujeron los costos de transporte en las autopistas, el tren convencional y la aviación. Se elevó el precio de gas con el fin de representar el incremento del impuesto a la gasolina para ayudar a financiar la línea ferroviaria de alta velocidad. Los analistas redujeron los costos de la vivienda en las regiones costeras de California e incrementaron los costos de la vivienda en la región central como una manera de representar el desplazamiento en los precios de la vivienda que se darían como resultado de la mayor facilidad de los desplazamientos entre comunidades hacia el interior del estado. Los analistas utilizaron también variables de la política detalladas para el consumidor y el gobierno para representar cambios en los patrones del gasto.

Energía y medio ambiente

Los economistas de la energía y el medio ambiente utilizan el modelo REMI para evaluar las implicaciones económicas de una amplia gama de políticas. La Agencia de Protección Ambiental de los Estados Unidos está utilizando el modelo para mostrar los beneficios económicos potenciales de los programas patrocinados por los estados para reducir la producción de gases invernadero.²⁶ Los pronósticos de mediano y largo plazo del REMI se utilizan para conducir modelos que estiman las emisiones, y que más tarde ayudan a identificar la necesidad de controles de la contaminación atmosférica para reducir estas emisiones. Otras aplicaciones ambientales del REMI incluyen los efectos económicos de la reducción de la lluvia ácida,²⁷ de programas de vehículos con bajas emisiones de gases,²⁸ de recuperación de lagos,²⁹ y de la restauración de los Everglades de Florida.³⁰

Las aplicaciones del REMI a la energía incluyen los efectos de cerrar tempranamente una planta de energía nuclear,³¹ una evaluación de la planta de producción de energía de Vermont,³² y los efectos de los programas para la eficiencia en la administración de la energía por parte de los consumidores.³³ En una evaluación de la desregulación del mercado de energía eléctrica en Wyoming, los analistas mostraron el efecto del incremento en las tarifas eléctricas sobre el estado. Los analistas predijeron que las tarifas eléctricas se incrementarían ya que las instalaciones del servicio público estarían en posibilidad de comercializar su energía de bajo costo relativo a una tarifa de mercado superior que prevalecería con la desregulación. Los cambios en los costos fueron implementados en el modelo REMI por separado para las variaciones del costo comercial, industrial

y para el consumo. El efecto de los incrementos en las tarifas de energía eléctrica para las empresas es la reducción en la competitividad empresarial y motiva una sustitución de la electricidad por un mayor uso de mano de obra y de capital. El incremento en las tarifas de energía eléctrica para el consumo resulta en menores ingresos reales, causando un pequeño efecto relativo de emigración. El desplazamiento del precio al consumo causa también una sustitución de energía eléctrica por otras categorías de consumo.³⁴

Aplicaciones del modelo REMI: ejemplos

Los analistas de políticas públicas utilizan el modelo REMI como un aspecto integral de la planeación y la evaluación. A menudo, la utilización del modelo cumple una orden de análisis de la organización. Por ejemplo, el Distrito para la Administración de la Calidad del Aire de la Costa Meridional, al sur de California, usa el modelo REMI para cumplir el requisito de evaluar los efectos socioeconómicos de todas las propuestas de regulación de la calidad atmosférica. En otro caso, la Contraloría del Estado de Texas efectuó una campaña con la plataforma para implementar análisis dinámicos de los impuestos para el estado. La Contraloría seleccionó el modelo REMI como medio para cumplir con este compromiso de campaña.

Varios departamentos de desarrollo económico del ámbito estatal utilizan el modelo REMI para medir los efectos de varios programas de incentivos para atraer empresas. Estos departamentos tienen a menudo autoridad para ofrecer beneficios fiscales como medio de alentar a las empresas a ubicarse en el estado. Sin embargo, estos incentivos son limitados de manera que las compañías tienen que alcanzar un umbral de creación de empleos para ser elegibles para obtener dichos beneficios.

Illinois, Michigan, Kentucky y Missouri son varios de los estados que aplican el análisis REMI como un componente estándar de la evaluación de sus paquetes de incentivos. Michigan, por ejemplo, reúne información sobre el producto y el empleo directos de la compañía y otros detalles específicos como tarifas salariales y planes de inversión. El estado evalúa la efectividad de los incentivos con base en las variaciones en el empleo total, los salarios, el producto bruto doméstico, los efectos fiscales en los gobiernos estatal y municipal, y el ingreso real per cápita. Con base en esta información, Michigan aprueba o niega el incentivo fiscal.

Estudio de caso: las plantas armadoras de automóviles de Mazda y Mitsubishi³⁵

Investigadores en la Universidad de Michigan iniciaron el uso del modelo REMI para la evaluación de los procesos para el otorgamiento de incentivos

fiscales en el ámbito estatal. La universidad, así como las agencias estatales, han desarrollado y refinado este tipo de procedimiento con el propósito de racionalizar el proceso de aplicación de los incentivos fiscales y proporcionar el máximo benéfico al público. Dos casos tempranos en Michigan fueron particularmente importantes para proporcionar un fundamento a la aplicación de métodos, y sirven como estudios de casos útiles para ilustrar el papel del análisis REMI en el proceso de diseño de la política.

A mediados de los ochentas, Mazda planeó construir una nueva planta para el armado de automóviles. Esta compañía se acercó a varios estados para pedir apoyo. Como parte de este proceso, investigadores de la Universidad de Michigan evaluaron tanto el aspecto de la localización como el impacto económico relacionados con la ubicación de la planta en Michigan. Este análisis se centra en la metodología del impacto económico, que fue central para determinar el paquete de incentivos del estado.

El estudio de Mazda partió del supuesto de que la planta de armado automotriz emplearía a 2500 trabajadores. Esta estimación se basó en plantas comparables en otras partes de Estados Unidos. El análisis incluyó también los gastos de inversión para modificar la planta existente de Flat Rock para armado de automóviles; se partió del supuesto de que este gasto de inversión tendría lugar en un periodo de tres años.

El resultado del análisis mostró la creación de 2357 empleos en total en el primer año de construcción, incrementándose a 13422 empleos durante el primer año de operación. Se predijo que la creación total de empleos comenzaría a estabilizarse entre los 12684 y los 12422 empleos los dos años de operación subsecuentes, respectivamente. El estudio también proporcionó diversas pruebas de sensibilidad, tales como una evaluación del efecto del salario sobre los incrementos en el empleo total.

El estado suministró un subsidio fiscal como parte de un esfuerzo para atraer a las instalaciones de Mazda. En un asunto relacionado, los investigadores de la Universidad de Michigan evaluaron los impactos económicos de la ubicación potencial de una planta de Mitsubishi en el estado. Mitsubishi solicitó incentivos a Michigan, Illinois y otros estados como parte de la elección de ubicación de la planta. En este caso, con base en los impactos económicos, los investigadores de Michigan sugirieron que el paquete de incentivos fiscales solicitado por Mitsubishi era demasiado grande. Como resultado de su recomendación, el estado se retiró de la licitación de pujas para la planta. En consecuencia, el estado salvó millones de dólares en incentivos fiscales que de otra manera hubieran sido entregados sin beneficio económico justificable. En definitiva, Mitsubishi ubicó la planta en Illinois.

Ejemplos de simulaciones

Para mostrar el desempeño del modelo REMI de Percepción de Políticas Públicas, hemos preparado muestras sencillas de simulaciones de políticas para varios ejemplos. En estos ejemplos, hemos hecho cambios únicamente en una o dos variables de las políticas con fines de ilustración. Casi todas las simulaciones de políticas realistas requieren que el usuario modifique diversas variables de la política para reflejar detalles de la que se propone.

Inmigración internacional

Esta simulación fue construida aumentando el nivel de emigración internacional para todos los grupos de edad. Como se incrementó el nivel de la migración internacional debido al supuesto de la política, hubo emigración neta de una parte de la fuerza laboral en respuesta al incremento en la oferta de mano de obra (y un declive subsiguiente en las oportunidades relativas de empleo y en la tarifa salarial relativa causado por la emigración internacional adicional).

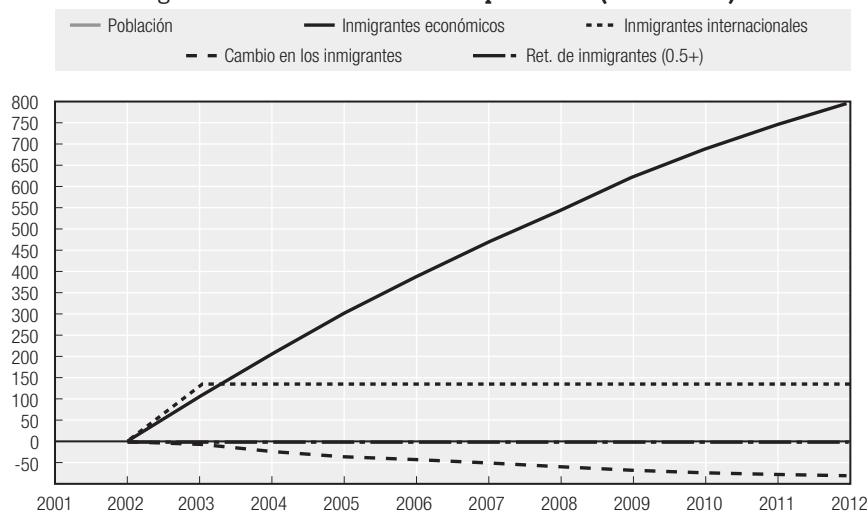
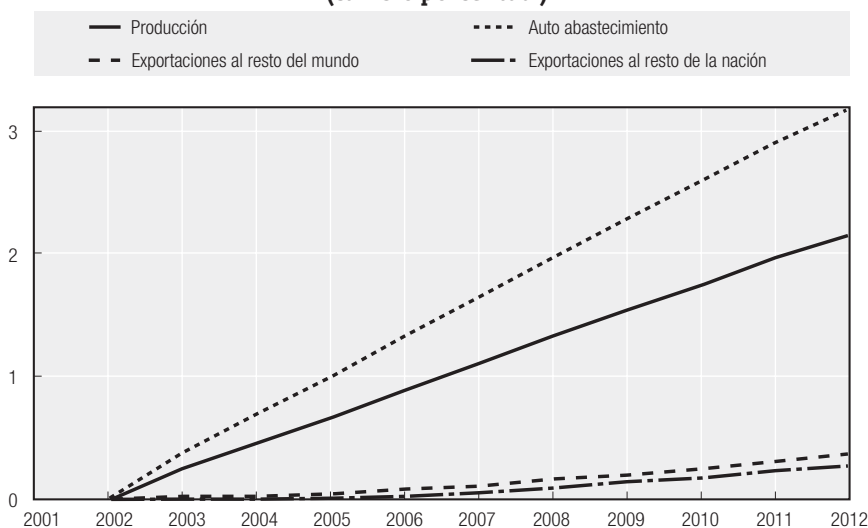
La fuerza laboral adicional suministrada por el incremento en la emigración internacional muestra que baja el costo de producción en un 0.091% para el 2010, ya que la disponibilidad de mano de obra causa que los salarios relativos sean un poco más bajos de lo que hubieran sido de otra forma. Las Figuras 6.4a y 6.4b muestran los valores de entrada capturados para las variables de la política, en este caso 133 mil por año. La Figura 6.5 muestra el incremento de la población para el cual hay tres componentes: el incremento en la emigración internacional causada por el supuesto del usuario, una caída compensatoria en la emigración económica en respuesta a la competencia de la fuerza laboral a la emigración internacional, y el cambio neto acumulado de la población en el horizonte de simulación. Como se muestra en la Figura 6.6, en esta simulación, el producto agregado de la ciudad de Nueva York crece en comparación con el escenario del punto de partida. Esto es inducido por un incremento en las exportaciones, el cual ocurre en la medida que las empresas se vuelven más competitivas debido a una mayor oferta de mano de obra y las consiguientes tarifas salariales relativas más bajas. La tabla de desempleo, Figura 6.7, muestra que la fuerza laboral crece más rápido que el empleo, dando como resultado un incremento en el índice de desempleo. Los índices de participación detallados muestran las variaciones para los diferentes grupos. Estos cambios son causados por las variaciones en el salario real y el índice de desempleo, así como las variaciones en la composición de las edades al interior de los grupos debido a la estructura de edades de los emigrantes internacionales.

Figura 6.4a Entradas de las variables de la política utilizadas

Variable	Detail	Units	Number	2001	2002	2003	2004	11
International Migration, All Ages, All Groups (Number)	All Ages	Thousands	040	0	0	133	133	11

Figura 6.4b Entradas de las variables de la política utilizadas

Variable	Detail	2005	2006	2007	2008	2009	2010	
International Migration, All Ages, All Groups (Number)	All Ages	133	133	133	133	133	133	133

Figura 6.5 **Determinantes de la población (diferencias)**Figura 6.6 **Componentes de la producción por el lado de la demanda (cambio porcentual)**

Incremento permanente del uno por ciento en la productividad total de los factores en 2003 correspondiente al Área Metropolitana de Rotterdam

Una simulación utilizando un modelo de tres regiones de los Países Bajos muestra la respuesta económica a un incremento en la productividad total de los factores (PTF). Se partió del supuesto de un incremento permanente del 1% de la PTF en el Área Metropolitana de Rotterdam a partir de 2003.

Figura 6.7 Tabla de desempleo

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Employment (% of employed)	0.000%	0.000%	0.212%	0.588%	0.861%	1.134%	1.407%	1.679%	1.950%	2.212%	2.473%	2.713%
Labor Force (% of labor force)	0.000%	0.000%	1.281%	2.282%	3.381%	4.289%	5.075%	5.822%	6.479%	7.106%	7.689%	8.238%
Unemployment rate	0.000%	0.000%	0.969%	1.694%	2.520%	3.155%	3.668%	4.143%	4.529%	4.894%	5.216%	5.525%
Participation rate	0.000%	0.000%	0.037%	0.035%	-0.034%	-0.037%	-0.211%	-0.325%	-0.446%	-0.576%	-0.692%	-0.816%
Total Male PTF (15-24)	0.000%	0.000%	0.276%	0.478%	0.505%	0.573%	0.598%	0.296%	0.189%	0.245%	0.262%	-1.002%
Total Female PTF (15-24)	0.000%	0.000%	0.602%	1.043%	1.402%	1.688%	1.796%	1.779%	1.680%	1.400%	1.332%	1.112%
Total Male PTF (25-64)	0.000%	0.000%	0.029%	-0.248%	-0.446%	-0.591%	-0.771%	-0.981%	-1.126%	-1.299%	-1.469%	-1.632%
Total Female PTF (25-64)	0.000%	0.000%	0.117%	-0.270%	-0.529%	-0.786%	-1.043%	-1.305%	-1.561%	-1.809%	-2.052%	-2.289%
Total Male PTF (65+)	0.000%	0.000%	0.478%	-1.073%	-1.739%	-2.520%	-3.246%	-4.004%	-4.718%	-5.437%	-6.164%	-6.796%
Total Female PTF (65+)	0.000%	0.000%	0.276%	0.039%	-1.174%	-1.746%	-2.348%	-2.974%	-3.616%	-4.248%	-4.879%	-5.524%
White NH Male PTF (15-24)	0.000%	0.000%	0.641%	1.498%	2.897%	3.982%	3.956%	3.152%	2.247%	1.364%	0.000%	0.271%
White NH Female PTF (15-24)	0.000%	0.000%	1.054%	1.898%	2.824%	3.351%	3.812%	4.184%	4.482%	4.782%	4.979%	4.262%
White NH Male PTF (25-64)	0.000%	0.000%	0.094%	-0.247%	-0.416%	-0.601%	-0.801%	-1.009%	-1.209%	-1.411%	-1.609%	-1.804%
White NH Female PTF (25-64)	0.000%	0.000%	0.134%	-0.249%	-0.609%	-0.891%	-1.195%	-1.489%	-1.785%	-2.074%	-2.366%	-2.669%
White NH Male PTF (65+)	0.000%	0.000%	0.532%	-1.371%	-2.130%	-2.923%	-3.718%	-4.504%	-5.213%	-5.891%	-6.567%	-7.044%
White NH Female PTF (65+)	0.000%	0.000%	0.044%	-0.152%	-0.486%	-0.934%	-1.482%	-2.060%	-2.700%	-3.380%	-4.002%	-4.506%
Black NH Male PTF (15-24)	0.000%	0.000%	0.312%	0.289%	0.502%	-0.084%	-0.299%	-0.689%	-1.066%	-1.528%	-1.940%	-2.286%
Black NH Female PTF (15-24)	0.000%	0.000%	0.674%	1.076%	1.302%	1.434%	1.489%	1.423%	1.286%	1.027%	0.000%	0.311%
Black NH Male PTF (25-64)	0.000%	0.000%	0.094%	-0.292%	-0.436%	-0.630%	-0.827%	-1.028%	-1.229%	-1.402%	-1.580%	-1.784%
Black NH Female PTF (25-64)	0.000%	0.000%	0.023%	-0.119%	-0.264%	-0.443%	-0.649%	-0.886%	-1.084%	-1.302%	-1.520%	-1.748%
Black NH Male PTF (65+)	0.000%	0.000%	0.564%	-1.258%	-2.175%	-3.076%	-3.922%	-4.609%	-5.126%	-5.602%	-6.070%	-6.502%
Black NH Female PTF (65+)	0.000%	0.000%	1.252%	-2.621%	-3.990%	-5.168%	-6.281%	-7.348%	-8.331%	-9.277%	-10.171%	-1.064%
Other NH Male PTF (15-24)	0.000%	0.000%	0.232%	-0.476%	-0.937%	-0.681%	-0.708%	-0.642%	-0.544%	-1.298%	-1.639%	-1.506%
Other NH Female PTF (15-24)	0.000%	0.000%	0.379%	0.764%	1.305%	1.689%	1.827%	1.824%	1.781%	1.693%	1.549%	1.397%
Other NH Male PTF (25-64)	0.000%	0.000%	0.154%	-0.323%	-0.482%	-0.629%	-0.791%	-0.959%	-1.154%	-1.384%	-1.624%	-1.771%
Other NH Female PTF (25-64)	0.000%	0.000%	0.449%	0.964%	1.432%	1.854%	2.251%	2.608%	2.933%	3.236%	3.524%	3.798%

La Figura 6.8 muestra la respuesta del producto agregado por la fuente de la demanda para la región del Área Metropolitana de Rotterdam. En la medida que la PTF incrementa y mejora la competitividad de las industrias de Rotterdam, sus ventas al resto de los Países Bajos y los mercados internacionales se incrementan. Como resultado del cambio único y permanente en 2003, el incremento en el autoabasto es mayor al 2% en el largo plazo, mientras que las exportaciones al resto del mundo y a las multirregiones crecen poco más del 1%.

La Figura 6.9 muestra el empleo por fuente de la demanda para el Área Metropolitana de Rotterdam. En el corto plazo, se da una caída en el empleo ya que la mayor productividad requiere menos trabajadores para el mismo nivel de producto. Sin embargo, el incremento en competitividad causado por los incrementos en PTF resulta en un incremento positivo en el empleo total después de 2 años. Esto es causado por el incremento en el empleo debido a las mayores exportaciones y también al proceso de ajuste del inventario de capital. En este proceso, las actividades de inversión y el empleo relacionado con la inversión se incrementan rápidamente a medida que la economía se expande. El incremento en la inversión disminuye gradualmente en cierto modo a medida que el inventario de capital crece a un nuevo nivel.

Figura 6.8 Componentes de la producción por el lado de la demanda (cambio porcentual)

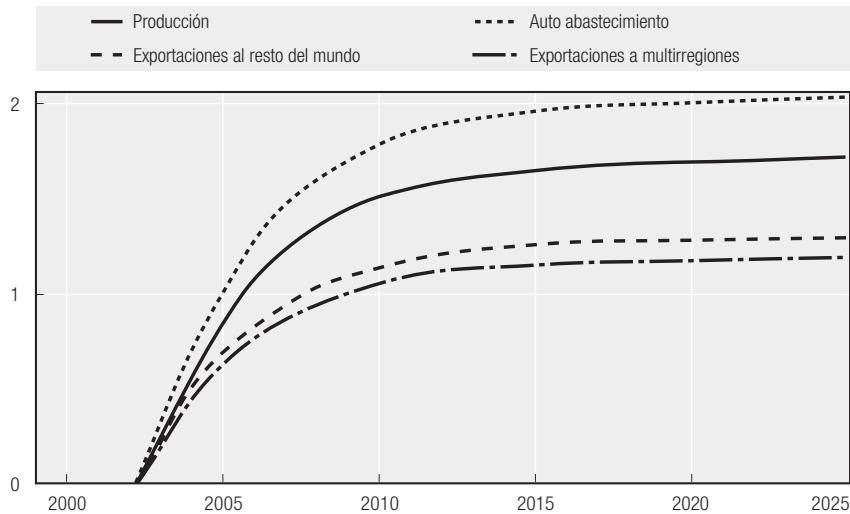


Figura 6.9 Componentes del empleo por el lado de la demanda (cambio porcentual)

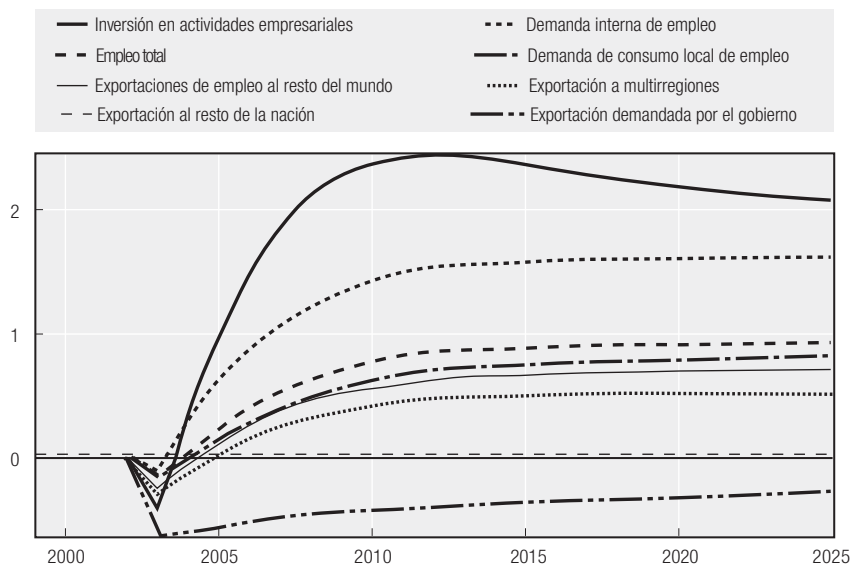
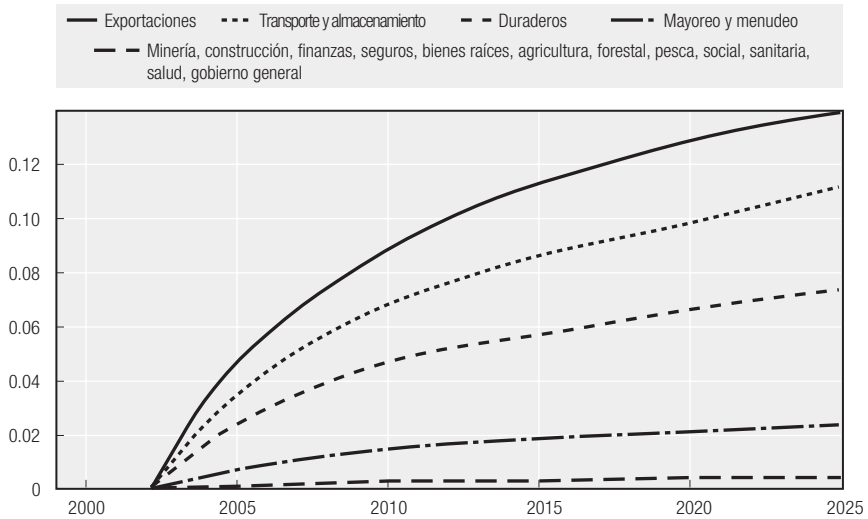


Figura 6.10 Exportaciones al resto del mundo (diferencias)



Los flujos comerciales entre las regiones en el modelo se muestran en la Figura 6.11. Para una industria especializada construida de madera, muebles, etcétera, esta tabla muestra que el Área Metropolitana de Rotterdam provee más de lo que son sus necesidades para esta industria, mientras que importa menos del resto del mundo. También hay un incremento en el comercio entre todas las regiones de los Países Bajos debido a la expansión de la economía nacional en la medida en que está centrada alrededor de la economía del Área Metropolitana de Rotterdam.

La Figura 6.12 resume las variaciones correspondientes al Área Metropolitana de Rotterdam. La mayor productividad en las industrias de Rotterdam conduce a una disminución en el costo de la producción, el precio de entrega y el índice de precios. La productividad mejorada conduce a un incremento en las exportaciones que orienta los incrementos en el empleo, el producto regional bruto y el ingreso personal.

Figura 6.11 Matriz del flujo comercial (diferencias)

2003		Wood, wood prod, lum, jewelry, games, toys		Type: Differences		
	Greater Rcd	Remaining	Rest of Na	Rest of Na	Rest of Wd	Output
Greater Rotterdam	0.0020681	0.00027004	0.00051905	0	0.00072996	0.0035871
Remaining South Holland	0.00039914	0.00026566	0.00014043	0	5.6744E-5	0.00086212
Rest of Netherlands	0.0012852	0.00028682	0.0018396	0	0.00016403	0.0035763
Rest of Nation	0	0	0	X	X	X
Rest of World	-0.00034183	0.0002867	0.00093746	X	X	X
Demand	0.0034106	0.0011094	0.0034361	X	X	X

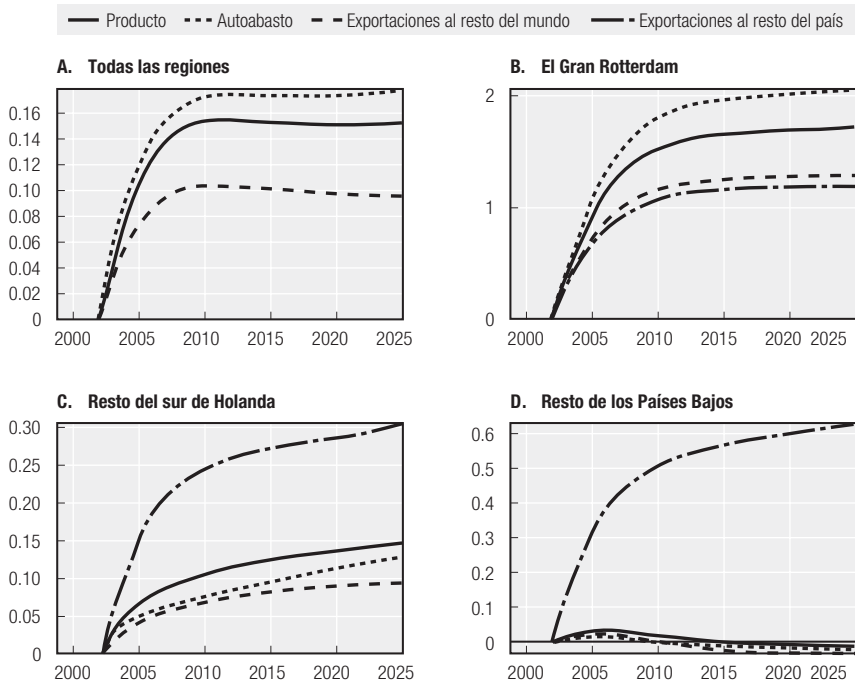
Figura 6.12 Tabla resumen (cambio porcentual)

Variable	2003	2004	2007	2010	2020
Employment (Thous)	+0.015%	+0.020%	+0.020%	+0.010%	-0.013%
GRP (1,000M 98eur)	+0.014%	+0.019%	+0.016%	+0.003%	-0.028%
Pers Inc (1,000M Nom eur)	+0.007%	+0.010%	+0.011%	+0.006%	-0.008%
PCE-Price Index (NL 1998=100)	-0.011%	-0.009%	-0.005%	-0.001%	+0.001%
Real Disp Pers Inc (1,000M 98eur)	+0.017%	+0.018%	+0.014%	+0.007%	-0.009%
Population (Thous)	+0.000%	+0.000%	+0.001%	+0.001%	-0.001%
Total Migrants	+0.053%	+0.049%	+0.030%	-0.005%	-0.064%
Labor Force	+0.002%	+0.003%	+0.006%	+0.005%	-0.004%
Demand (1,000M 98eur)	+0.018%	+0.023%	+0.025%	+0.016%	-0.007%
Output (1,000M 98eur)	+0.017%	+0.024%	+0.028%	+0.017%	-0.010%
Delivered Price	-0.013%	-0.011%	-0.006%	-0.002%	+0.001%
Rel Cost of Production	-0.007%	-0.005%	+0.000%	+0.004%	+0.007%
Labor Intensity	0.000%	0.000%	0.000%	0.000%	0.000%
Labor Access Index	0.000%	0.000%	+0.001%	+0.002%	+0.004%
Indust Mix Index	0.000%	0.000%	0.000%	0.000%	0.000%
Reg Pur Coeff (SS over Dem)	-0.002%	-0.004%	-0.010%	-0.016%	-0.027%
Imports (1,000M 98eur)	+0.022%	+0.031%	+0.045%	+0.047%	+0.042%
Self Supply (1,000M 98eur)	+0.016%	+0.020%	+0.015%	0.000%	-0.034%
Exports to Multiregions (1,000M 98eur)	+0.103%	+0.210%	+0.414%	+0.507%	+0.601%
Exports to Rest of Nation (1,000M 98eur)	N/A	N/A	N/A	N/A	N/A
Exp to Rest of World (1,000M 98eur)	+0.008%	+0.012%	+0.014%	+0.005%	-0.020%
Wage Rate (Thous Nameur)	+0.001%	+0.002%	+0.003%	+0.002%	-0.004%

Las Figuras 6.13a-6.13d muestran el cambio porcentual en los componentes del producto total por la fuente de la demanda para todas las regiones en los Países Bajos, la región del Área Metropolitana de Rotterdam, el Remanente de Holanda Meridional y el Resto de los Países Bajos. Aunque la principal expansión en las participaciones de mercado con el resto del mundo se da en las exportaciones,

el incremento en el autoabasto es un incremento más significativo debido a la expansión de la economía local. Lo anterior se muestra con las variaciones en el autoabasto y las exportaciones al resto del mundo. La Figura 6.13c muestra un incremento en las exportaciones a otras partes de los Países Bajos provenientes del Remanente de Holanda Meridional. La Figura 6.13d muestra las variaciones del Resto de los Países Bajos. Ambas Figuras, 6.13c y 6.13d, muestran que aunque el incremento en la productividad ocurre en el Área Metropolitana de Rotterdam, la expansión de la economía del Área Metropolitana de Rotterdam lleva a que algunas de las necesidades de importación de dicha zona sean cubiertas por el incremento en las exportaciones multirregionales desde otras regiones en el modelo de los Países Bajos.

Figura 6.13 **Componentes del producto total por la fuente de la demanda (cambios porcentuales)**



Reducción en el tiempo del recorrido debido a una mejora en la infraestructura del transporte

Por último, presentamos una simulación simplificada con la versión europea del modelo REMI de Demostración de la Percepción de las Políticas Públicas. Este

modelo es de una sola zona para una pequeña área del Reino Unido. Hacemos esta simulación para ilustrar que el modelo REMI de Percepción de las Políticas nos hace posible probar y evaluar los modelos y de este modo mostrar todos los detalles que documentan y explican los resultados.

En este caso partimos del supuesto de que la intervención es una inversión en la infraestructura del transporte que disminuirá el tiempo de transporte interno en la región de la demostración en un 10% para el 50% de las millas recorridas por los vehículos. Simplificamos el análisis ingresando una reducción inmediata en los costos del transporte y la accesibilidad del 5% en el primer año y todos los subsecuentes. La información para hacer este tipo de estimación provendría directamente de un análisis costo-beneficio en la propuesta para la inversión o de un análisis de la ubicación y del alcance del proyecto. Podría incluir el uso medido y los ahorros de tiempo si la inversión se ha hecho en su totalidad.

Las variaciones resultantes a partir de los valores de control, tal como los imprime el modelo REMI de Percepción de Políticas, se muestran en la Figura 6.14 y la Tabla 6.1 como diferencias, y en la Tablas 6.2 y 6.3 como cambios porcentuales (véanse la Figura 6.14 y las Tablas 6.1, 6.2 y 6.3 a continuación).

Figura 6.14 **Muestra de las simulaciones del modelo de demostración**

Reducción de los costos del transporte interno y del acceso en un 5%

Organización analítica (usar la opción de la gráfica):

Gráfica 4: Componentes del producto total por el origen de la demanda

Diferencias comparadas con el REMI de control estándar de regulación



Tabla 6.1 REMI de control estándar de regulación
Reducción de los costos del transporte interno y del acceso en un 5%
Gráfica 4: Componentes del producto total por el origen de la demanda — diferencias

Variable	1998	2000	2010	2025
Producto total (1 000 M 97 £)	0.19170	0.24660	0.35970	0.3924
Autoabasto (1 000 M 97 £)	0.16910	0.18870	0.23530	0.2525
Exp. a multirregiones (1 000 M 97 £)	0	0	0	0
Exp. al resto del país (1 000 M 97 £)	0.01009	0.02580	0.04763	0.0378
Exp. al resto del mundo (1 000 M 97 £)	0.01246	0.03203	0.07679	0.1021

En la Figura 6.14 y las Tablas 6.2 a 6.3, tres componentes del producto total varían (las exportaciones multirregión permanecen sin cambios porque se trata de un modelo de una sola región). Las exportaciones al resto del mundo y al resto del país crecen debido a que el costo de producción decrece en relación con su valor en el año de calibración.

El autoabasto (en la Figura 6.14 y en todas las tablas) crece porque la participación de la zona de demostración abastecida localmente se incrementa (el coeficiente de compras regionales, Tabla 6.3) y la demanda se incrementa (Tabla 6.3). La participación local aumenta porque el costo relativo de producción disminuye (Tabla 6.3). La demanda crece porque el consumo (no mostrado) se incrementa debido al crecimiento en los incrementos en el ingreso real disponible (Tabla 6.3) y al incremento de la participación regional abastecida localmente (el Coeficiente de Compras Regionales). La demanda también se incrementa debido al crecimiento en la inversión (no mostrado) y al gasto gubernamental local (no mostrado). Los incrementos del ingreso real disponible se dan debido a que el empleo sube (Tabla 6.3) y los precios bajan (Tabla 6.3).

Tabla 6.2 REMI de control estándar de regulación
Reducción de los costos del transporte interno y del acceso en un 5%
Gráfica 4: Componentes del producto total por el origen
de la demanda — cambios porcentuales

Variable	1998	2000	2010	2025
Producto total (1 000 M 97 £)	0.74%	0.93%	1.26%	1.39%
Autoabasto (1 000 M 97 £)	2.10%	2.26%	2.49%	2.60%
Exp. a multirregiones (1 000 M 97 £)	n.a.	n.a.	n.a.	n.a.
Exp. al resto del país (1 000 M 97 £)	0.09%	0.23%	0.45%	0.47%
Exp. al resto del mundo (1 000 M 97 £)	0.18%	0.45%	0.89%	0.98%

Tabla 6.3 REMI de control estándar de regulación
Reducción de los costos del transporte interno y del acceso en un 5%
mayores cambios porcentuales

Variable	1998	2000	2010	2025
Empleo (miles)	1.23%	1.39%	1.61%	1.69%
Producto Regional Bruto (1 000 M 97 £)	0.83%	1.04%	1.40%	1.54%
Ingreso personal (1 000 M Nominal £)	0.47%	0.59%	0.81%	0.96%
Índice de precios personal (UK 1997 = 100)	-0.54%	-0.54%	-0.55%	-0.56%
Ingreso personal real disponible (1 000 M 97 £)	0.83%	0.92%	1.11%	1.30%
Población (miles)	0.01%	0.04%	0.21%	0.49%
Migración económica	1.56%	1.58%	2.10%	2.29%
Migración total	1.56%	1.58%	2.10%	2.29%
Fuerza laboral	0.17%	0.34%	0.71%	1.00%
Demanda (1 000 M 97 £)	0.74%	0.90%	1.11%	1.21%
Producto (1 000 M 97 £)	0.74%	0.93%	1.26%	1.39%
Precio de entrega	-0.48%	-0.48%	-0.49%	-0.50%
Costo real de producción	-0.26%	-0.26%	-0.27%	-0.28%
Intensidad de la mano de obra	0.00%	-0.01%	-0.05%	-0.09%
Índice de acceso a la mano de obra	0.01%	0.03%	0.06%	0.07%
Índice industrial combinado	0.00%	0.00%	0.00%	0.00%
Coefficiente de compras regionales (SS/Demanda)	1.35%	1.35%	1.36%	1.38%
Importaciones (1 000 M 97 £)	-0.17%	-0.02%	2.49%	2.60%
Autoabasto (1 000 M 97 £)	2.10%	2.26%	2.49%	2.60%
Exportaciones multirregionales (1 000 M 97 £)	n.a.	n.a.	n.a.	n.a.
Exportaciones al resto del país (1 000 M 97 £)	0.09%	0.23%	0.45%	0.47%
Exportaciones al resto del mundo (1 000 M 97 £)	0.18%	0.45%	0.89%	0.98%
Tarifa salarial (Miles Nominal £)	-0.13%	-0.03%	0.15%	0.15%

La Tabla 6.3 muestra que la tarifa salarial nominal promedio disminuye ligeramente el primer año, aunque la tarifa salarial en cada industria aumenta (no se muestra). Esto sucede porque las industrias con bajos salarios (servicios, etcétera) incrementan el empleo a una tasa mayor que las industrias con altos salarios. El incremento en el índice de acceso a la mano de obra, que incrementa la productividad, se debe a la economía en expansión y al número de empleados. Obsérvese también que el Producto Bruto Doméstico per cápita incrementa en el corto y largo plazos (Producto Regional Bruto y Población, Tabla 6.3).

Además de la información mostrada, hay varios miles de otras variables disponibles. Las tablas y las gráficas incluyen información sobre todos los conceptos utilizados por industria y por grupos de edad. Utilizando la información mostrada, así como la información de otras gráficas y tablas del modelo REMI se puede documentar plenamente y explicar los efectos de la inversión en fondos para la estructura.

Recomendaciones para las políticas públicas

Las autoridades locales y centrales utilizan los modelos económicos para tomar decisiones más efectivas. Las autoridades gubernamentales tienen recursos limitados para aplicarlos al desarrollo económico, por lo tanto los modelos económicos y otras herramientas analíticas son importantes puesto que proporcionan un medio para evaluar usos alternativos de los recursos.

Sugerimos que los tomadores de decisiones utilicen modelos que posean las seis siguientes características clave: 1) especificación de las condiciones locales, 2) fundamentos teóricos y estructurales sólidos, 3) integración de los métodos de equilibrio general, insumo-producto, econométrico y de geografía económica, 4) un extenso conjunto tanto de variables de entrada como de variables de respuestas, 5) resultados año por año, 6) una utilización probada para un amplia variedad de proyectos en diferentes regiones.

La calibración del modelo para una economía específica debería incorporar datos locales. Sin embargo, los parámetros econométricos del modelo deberían basarse en un amplio conjunto de datos transversales y de series de tiempo. El conjunto de datos mayor debería ser utilizado con fines de estimación, ya que las estimaciones estadísticas robustas para un modelo estructural bien definido puede que no sean posibles utilizando datos de series de tiempo para una sola región.

El análisis de la política debería ser claro y completamente defendible. Un modelo que se fundamenta en bases teóricas y estructurales sólidas incorporará relaciones causa-efecto bien definidas en la economía. Las bases para las recomendaciones de las políticas derivadas de un modelo de este tipo serán totalmente transparentes y los diseñadores de las políticas estarán plenamente informados.

Un enfoque de modelo integral reúne las ventajas de diferentes metodologías en un sistema comprensivo y consistente. Las estructuras insumo-producto son importantes para escudriñar las relaciones entre industrias en la economía. Las respuestas del equilibrio general captan importantes reacciones de largo plazo a las señales de precio, costo y salario. Las técnicas econométricas validan las bases empíricas de un modelo, y los nuevos métodos de geografía económica explican que las economías de aglomeración son significativas para la comprensión de la dinámica del desarrollo regional.

Es necesario un marco de modelo exhaustivo, que incluya variables de política detalladas para todas las partes de la economía, a fin de representar adecuadamente los tipos de políticas que pueden ser propuestas para el desarrollo económico u otros propósitos. Los modelos que se desarrollan con un solo propósito o aplicación pueden ser bastante limitantes. Por ejemplo, un plan de desarrollo económico puede incluir un elemento de capacitación del

trabajador; por ello un modelo debería incorporar variables de políticas que permitan al usuario modificar la productividad de la mano de obra para poder captar completamente los efectos de la propuesta.

Tanto para propósitos de planeación como de análisis, es crucial un modelo dinámico con resultados año por año porque brinda una clara comprensión de las implicaciones del programa en el tiempo. Un pronóstico económico de mediano y largo plazos es invaluable para la determinación de la necesidad de carreteras, aeropuertos, plantas generadoras de energía, escuelas y otras inversiones. Muchas políticas también tendrán importantes implicaciones de largo plazo; por ejemplo, muchos de los beneficios económicos de nuevas inversiones para el transporte se dan décadas después de la construcción inicial. Los analistas deben comprender también la temporalidad de los efectos con el propósito de evaluar la factibilidad total del proyecto.

El desempeño de un modelo económico se debería validar idealmente a través de su aplicación a numerosos estudios para diversas economías. Aunque se puede construir y probar un modelo en su desarrollo inicial, su comportamiento en muchos estudios del mundo real es vital como parte del proceso de desarrollo de un modelo. Los diseñadores de modelos pueden mejorar y refinar continuamente un modelo económico a partir de las experiencias de los usuarios del modelo.

Notas

1. Véase Fan, Wei, Treyz y Treyz (2000).
2. Véase Treyz, Frederick y George Treyz. "The REMI Economic Geography Forecasting and Policy Analysis Model". 1 de agosto de 2001.
3. Véase Treyz, George y Lisa Petraglia (2001).
4. Véase Rickman, Dan S., Gang Shao y George I. Treyz (1992).
5. Véase Greenwood *et al.* (1991).
6. Véase REMI Staff (2002). "REMI Policy Insight, Model Documentation, European Version 5.1", Regional Economic Models, Inc., pp. 46-52.
7. Prof. Dr. Michael Wegener, Institute of Spatial Planning (IRPUD), August-Schmidt-Str.6, 44221 Dortmund, Alemania
8. Véase Weisbrod *et al.* (2001).
9. Véase Systematic Solutions, Inc. (1992).
10. Véase Minnesota Racing Commission (1991).
11. Véase Lenk, Franklin (1990).
12. Véase Blois, Tara *et al.* (1995); Deloitte y Touche LLP (1995); y Sims, Richard (1994).
13. Véase Kentucky Legislative Research Committee (1986).

14. Véase Fulton, Grimes y Baum (1984).
15. Véase Boston Redevelopment Authority (2000) y Connecticut Center for Economic Analysis (1992).
16. Véase Connecticut Center for Economic Analysis (1993) y Deller, Steven C. et al. (1992).
17. Véase Warner et al. (1996).
18. Véase Bonardelli, Mark A. (1995).
19. Véase Bartlett et al. (2001).
20. Véase Otto, D. y M. Lipsman (1993).
21. Véase University of Connecticut (1994).
22. Véase Nexus Associates, Inc. (1995).
23. Véase Harkenrider, Grez (1999).
24. Véase Wilbur Smith Associates (1992).
25. Véase Economic Research Associates (1996).
26. Véase Smith, Anne E. et al. (1997).
27. Véase Myers, J.G., C.A. Pasurka, J., y T. Veselka (1987).
28. Véase ICF, Inc. (1993).
29. Véase Duncombe, William et al. (1997).
30. Véase Weiskoff, Richard (2000) y Weiskoff, Richard (2002).
31. Véase Silkman, Richard (1987).
32. Véase Vermont Department of Public Service (1991).
33. Véase Hickman, James E. (1995) y Weisbrod et al. (1995).
34. Véase Wyoming Public Service Commission (1996).
35. Véase Baum, A.L., G. Fulton y D.R.Grimes (1984).

Referencias bibliográficas

- BAUM, A.L., G. FULTON y D.R. GRIMES, "Industrial Location Decisions and Their Impact on the Michigan Economy: The Mazda Automobile Assembly Case", 1984.
- BLOIS, Tara, Steven R. CUNNINGHAM y William F. LOTT, "The Bridgeport Casino Proposals". Connecticut Center for Economic Analysis, 42 páginas; octubre, 1995.
- BONARDELLI, Mark A., "Analyzing the Impact of Lost Coal Sales Using the Illinois REMI Model". *International Journal of Public Administration*, vol. 18, núm. 1, pp. 101-118; 1995.
- BOSTON REDEVELOPMENT AUTHORITY POLICY DEVELOPMENT AND RESEARCH DEPARTMENT, "An Economic Analysis of Fenway Park and the New Fenway Park Proposal (2000)"; 13 de abril de 2000.
- CHRISTOPHER, JR., Chris G., CHENGFENG LOU y George I. TREYZ, "Regional Labor Force Participation Rates"; 3 de marzo de 1996.

- CONNECTICUT CENTER FOR ECONOMIC ANALYSIS, "New England Patriots Franchise Acquisition: An Economic Impact Study", 41 páginas; 12 de noviembre de 1992.
- DELOITTE y TOUCHE LLP, "Economic Impacts of Casino Gaming on the State of Michigan", 114 páginas; febrero 1995.
- DUNCOMBE, William, Shannon FELT, James R. FOLLAIN y Bernard JUMP, Jr., "The Economic and Fiscal Impact of Lake Remediation on Onondaga County". Center for Policy Research; Metropolitan Studies Program Series Occasional Paper núm. 186; junio 1997.
- ECONOMIC RESEARCH ASSOCIATES, "Economic Impact and Benefit/Cost of High Speed Rail for California (final report)"; 1996, 154 páginas.
- FAN, Wei, Frederick TREYZ y George TREYZ, "An Evolutionary New Economic Geography Model", *Journal of Regional Science*, vol. 40(4), 671-695, 2000.
- FUJITA, Masahisa, Paul KRUGMAN y Anthony J. VENABLES, *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, Massachusetts: MIT Press, 1999.
- FULTON, George, comunicación personal; 1 de noviembre de 2002.
- FULTON, G., D.R. GRIMES y A.L. BAUM, "Industrial Location Decisions and Their Impact on the Michigan Economy: The Mazda Automobile Assembly Case". 58 páginas; noviembre 1984.
- GREENWOOD, Michael J., Gary L. HUNT, Dan S. RICKMAN y George I. TREYZ, "Migration, Regional Equilibrium, and the Estimation of Compensating Differentials". *American Economic Review*; 1991.
- HARKENRIDER, Greg, "Prison Study Economic Report", Commonwealth of Kentucky. 17 páginas; septiembre 1999.
- HICKMAN, James E., "Economic Opportunities Through Energy Efficiency: An Alternative Analysis". 12 páginas; enero 1995.
- ICF, INC., "Economic Effects of Adopting a Low Emission Vehicle Program in Maryland". Preparado para la American Lung Association of Maryland, Inc.; Febrero 1993.
- KENTUCKY LEGISLATIVE RESEARCH COMMITTEE, "Economic and Fiscal Effects of the Toyota Auto Facility on the Kentucky Economy". 13 páginas; 1986.
- LENK, Franklin, "The Economic Impact of Expanding Bartle Hall and Building an 800-Room Hotel". 26 páginas; 6 de febrero de 1990.
- OTTO, D. y M. LIPSMAN, "Economic Impacts of the 1993 Iowa Floods". 6 páginas; 1993.
- MINNESOTA RACING COMMISSION, "The Economic Impact of the Horse Racing and Breeding Industry on the State of Minnesota". 9 páginas; abril 1991.
- MYERS, J.G., C.A. PASURKA, Jr. y T. VESELKA, "An Input-Output Simulation of the Impact of Acid Rain Legislation on Illinois". 37 páginas; agosto 1987.
- NEXUS ASSOCIATES, INC., "Executive Summary: The Impact of Tufts University School of Veterinary Medicine on the Massachusetts Economy". 29 páginas; octubre 1995.
- REMI STAFF, 2002, "REMI Policy Insight, Model Documentation, European Version 5.1", Regional Economic Models, Inc.

- REMI STAFF 2002, "REMI Policy Insight, User Guide, Version 5.1", Regional Economic Models, Inc.
- RICKMAN, Dan S., Gang SHAO y George I. TREYZ, "Multi-regional Stock Adjustment Equations of Residential and Non-residential Investment in Structures", *Journal of Regional Science*, 1993.
- RICKMAN, Dan S., Gang SHAO y George I. TREYZ, "The REMI Economic-Demographic Forecasting and Simulation Model", *International Regional Science Review*, 1992.
- SILKMAN, Richard, "The Effects of a Mandatory Early Shutdown of Maine Yankee". 38 páginas con 52 páginas de apéndices; septiembre 1987.
- SIMS, Richard, "Economic Impact Analysis: The Lucky Spa Casino". Office of Economic and Tax Policy, Bureau of Legislative Research. 5 páginas; 1994.
- SMITH, Anne E. et al., "Costs, Economic Impacts, and Benefits of EPA's Ozone and Particulate Standards". Reason Public Policy Institute/Decision Focus Inc., Policy Study 226. 114 páginas; junio 1997.
- SYSTEMATIC SOLUTIONS, INC., "REMI Interface with Energy 2020". 18 páginas; 1992.
- TREYZ, Frederick y Jim BUMGARDNER, "Monopolistic Competition Estimates of Interregional Trade Flows in Services". *Regional Cohesion and Competition in the Age of Globalization*, 2000.
- TREYZ, Frederick y George TREYZ, "The REMI Economic Geography Forecasting and Policy Analysis Model". 1 de agosto de 2001.
- TREYZ, Frederick y George I. TREYZ, "The REMI Multi-regional US Policy Analysis Model". Reunión Anual Norteamericana de la Regional Science Association, 1997.
- TREYZ, George y Lisa PETRAGLIA, "Consumption Equations for a Multi-regional Forecasting and Policy Analysis Model". *Regional Science Perspectives in Economic Analysis*. Comps. Michael Lahr y Ronald Miller. Elsevier Science B.V., 287-300, 2001.
- TREYZ, George I., "Policy Analysis Applications of REMI Economic Forecasting and Simulation Models". *International Journal of Public Administration*, 1995.
- UNIVERSITY OF CONNECTICUT, "1995 World Special Olympics — Economic Impact Analysis". 11 páginas; abril 1994.
- VERMONT DEPARTMENT OF PUBLIC SERVICE, "Vermont Comprehensive Energy Plan". 203 páginas; enero 1991.
- WARNER, K.E., G.A. FULTON, P. NICOLAS y D. GRIMES, "Employment Implications of Declining Tobacco Product Sales for the Regional Economies of the United States". *The Journal of the American Medical Association*. vol. 275, núm. 16, pp. 1241-46; 24 de abril de 1996.
- WEISBROD, Glen de Hagler Bailly Consulting, INC., Karen POLENSKE y Teresa LYNCH del MIT y Xianuan LIN de Boston University, "Final Report: The Economic Impact of Energy Efficiency Programs and Renewable Power for Iowa".
- WEISBROD, Glen, Donald VARY y George TREYZ, "Project 2-21 — Final Report: Economic Implications of Congestion". *Transportation Research Board, National Cooperative*

Highway Research Program Report 463. Washington, DC: National Academy Press, 2001.

WEISKOFF, Richard, "Missing Pieces in Ecosystem Restoration: The Case of the Florida Everglades". *Economic Systems Research*, vol. 12, núm. 3, 2000.

WEISKOFF, Richard, "The Economics of Everglades Restoration: A Tale to Two Models". School of International Studies, University of Miami. 16 páginas; 2002.

WILBUR SMITH ASSOCIATES, "Executive Summary: US Highway 20 Corridor Development Study". 18 páginas; 1992.

WYOMING PUBLIC SERVICE COMMISSION, "Electric Utility Industry Restructuring Issues". Resumen de seis subcomités de las partes interesadas en Wyoming. 61 páginas; 12 de noviembre de 1996.

Capítulo 7

Comentarios sobre la ponencia de Frederick y George Treyz y el taller “Políticas de análisis para el desarrollo local con modelos de pronóstico”

por
Robert Wilson
Directivo del Instituto de Investigación del Empleo
Warwick University, Reino Unido

La presentación

El Taller comenzó con una breve introducción al modelo REMI, en la que se planteó la funcionalidad del modelo en términos de sus posibles aplicaciones. Posteriormente, el Dr. Treyz describió los elementos clave del modelo:

- La estructura insumo-producto (*input-output*)
- Elementos generales de equilibrio, con reacciones de largo plazo a precios y salarios.
- El componente econométrico con base en modelos de series de tiempo para incorporar efectos dinámicos.
- Una nueva geografía económica que incluye efectos de aglomeración y agrupación.

En seguida, se tomaron como ejemplo algunos insumos que podían servir para construir el modelo, así como los potenciales resultados de este ejercicio, enfocándose en la demanda de bienes y servicios, los efectos en precios y salarios, y los efectos demográficos y en los mercados laborales.

Posteriormente, se analizaron —para algunos países, entre ellos Estados Unidos— propuestas de políticas económicas que podrían derivarse de los resultados del modelo, con ejemplos de diferentes contextos.

Se discutieron las ventajas y las desventajas de adoptar enfoques basados en un modelo. Los puntos relevantes de la discusión fueron los siguientes:

- La importancia del análisis macroeconómico para la toma de decisiones de las políticas; los estudios microeconómicos solamente constituyen una parte del proceso de evaluación.
- Un modelo macro cuantitativo genera un escenario común que permite comparar los efectos fiscales y de empleo de diversas intervenciones o políticas.
- Un enfoque global nos permite tomar en cuenta los efectos de desplazamiento entre fronteras geográficas.

- Un modelo general permite la incorporación de conceptos teóricos que se apoyan en pruebas empíricas de diversas disciplinas.
- El modelo facilita la transparencia en términos de supuestos e implicaciones.
- El componente dinámico del modelo es esencial para evaluar la ocurrencia de los efectos en el tiempo.

Se recalcó la necesidad de que el uso de estos métodos sea un proceso, con retroalimentación de los usuarios y con las subsiguientes modificaciones que influyan en los resultados finales. El análisis no debe ser tan solo un ejercicio aislado.

Si se lleva a cabo adecuadamente, puede dotar a los usuarios de los medios necesarios y asegurar su inclusión y posesión de los logros y resultados.

Modelos macro-económicos alternos

Existen otros modelos macro-económicos que pueden utilizarse para evaluar las intervenciones en el mercado laboral. Por ejemplo, en el Reino Unido, el Instituto de Investigación del Empleo (IER, por sus siglas en inglés) y Cambridge Econometrics (CE) desarrollaron el Modelo de Pronósticos para Economías Locales (LEFM, por sus siglas en inglés). Al igual que el REMI, el modelo LEFM supone patrones comunes de comportamiento y vinculaciones técnicas a nivel nacional, regional y local, si bien reconoce las diferencias estructurales únicas de cada área geográfica.

El modelo LEFM es un programa de *software* específico que permite hacer análisis y pronósticos de la economía local en el Reino Unido. Desde su inicio en 1993, ha sido implementado en cientos de lugares del Reino Unido y en algunas cuantos de Europa. Su principal función es informar a los diseñadores de políticas públicas y a los analistas sobre las tendencias fundamentales económicas y del mercado laboral a nivel local.

El modelo LEFM ha sido diseñado para satisfacer los siguientes criterios:

- Un medio eficiente para generar proyecciones económicas locales, hechas a la medida, utilizando al máximo la información disponible a nivel local, regional y nacional.
- Permitir fácilmente la actualización de información y la recopilación de análisis pasados.
- Un método de análisis riguroso y transparente que genere resultados verificables a partir de los supuestos.
- Un fácil enlace a pronósticos oficiales y actualizados, a nivel regional y nacional.
- Permitir, de manera explícita, la inclusión de puntos de vista e información local.

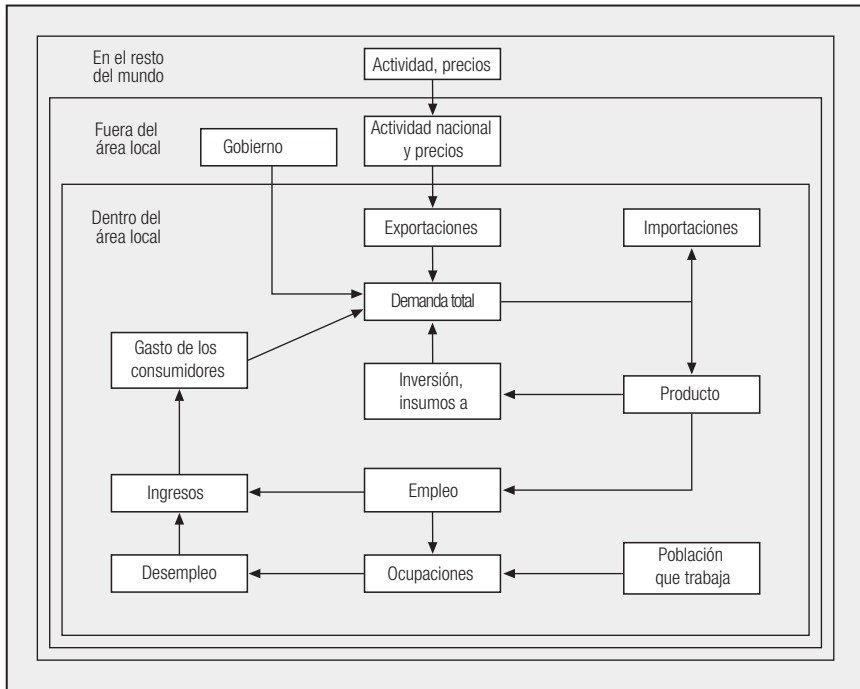
- Incluir información detallada a nivel sectorial; a fin de identificar las proyecciones con las principales empresas locales.
- Incluir información detallada del mercado laboral (sector, ocupación, género).
- Capacidad de implementar análisis de sensibilidad y cambios de escenarios.
- Capacidad para llevar a cabo estudios de impacto (como lo sería, por ejemplo, la apertura o el cierre de una gran empresa).
- Un acceso fácil a los resultados para su evaluación; transferencia de información y de gráficas a otros programas para redactar informes, hacer presentaciones, etcétera.

En el modelo LEFM la información sobre el empleo ocupa un papel central debido a que para dicha variable se dispone de la información más detallada y confiable a nivel local. Casi todos los indicadores económicos a nivel local se obtienen al combinar la información del indicador correspondiente a nivel regional y la cuota de participación de empleos a nivel local en un contexto regional o poblacional.

La estructura básica del modelo LEFM comparte algunos aspectos con el modelo REMI (Ver Figura 7.1). Se basa en una tabla Leontieff de insumos-productos. Ofrece una solución dinámica utilizando datos anuales que se remontan a 1981 y proyecciones que van hasta el 2015. En contraste con el modelo REMI, sin embargo, el LEFM se enfoca primordialmente en la demanda. Distingue entre cincuenta distintas actividades de empleo y seis categorías de estado-género. Adicionalmente, los resultados ocupacionales son desglosados en dos categorías separadas para hombres y mujeres. Comprende una serie de relaciones técnicas y de comportamiento a nivel local que a su vez reflejan las relaciones a nivel regional y nacional (valores parámetro, etcétera). Sus principales elementos son:

- Demanda de los consumidores
- Demanda de inversión
- Gastos de gobierno
- Exportaciones e importaciones
- Demanda intermedia
- Demanda total de bienes (Q)
- Producción industrial bruta (Y)
- Valor o
- Empleo Industrial (YEO)

Figura 7.1 LEFM — Modelo de Pronósticos para Economías Locales



- Empleo por tipo y género
- Empleo por profesión
- Oferta laboral, etcétera.
- Desempleo
- Ingresos

El modelo LEFM es calibrado de tal manera que el empleo a nivel local sigue los patrones históricos, mientras que los pronósticos siguen las tendencias nacionales o regionales. Lo anterior se basa en un simple análisis econométrico. La calibración se logra modificando los niveles locales de producción bruta mediante enmiendas a la tasa de participación en las exportaciones. Este método asegura que el modelo pueda escrutarse la información y las observaciones del empleo y que el pronóstico inicial sea “sensato”. Ciertos ajustes aseguran que el modelo verifique a su vez los datos de desempleo y que el pronóstico básico inicial cambie de acuerdo con las tendencias generales de cada región. Los detalles completos del modelo se pueden encontrar en Wilson *et al.* (1995).

Discusión inicial: importancia y validación estadística

Se puso en tela de juicio la solidez estadística de los resultados del método REMI, así como la aplicabilidad de parámetros estadounidenses al contexto europeo. Lo anterior forma parte de un debate más general en torno a los resultados de modelos basados en experiencias históricas particulares y su aplicabilidad o no-aplicabilidad a otros contextos y periodos.

A este respecto, el Dr. Treyz, aseguró que se han tomado importantes medidas para asegurar la robustez estadística de los parámetros estimados; sin embargo reconoció que en el caso de los pronósticos, es mucho más difícil formular declaraciones sobre su precisión estadística. Más bien, debemos enfatizar la generación de resultados razonables y sensatos, examinar la diferencia entre las distintas políticas y el contraste entre aplicarlas o no aplicarlas.

Por otro lado, se generó una prolongada discusión sobre los problemas de aplicar el modelo REMI al contexto europeo —dadas las diferencias económicas y sociales. Además del problema de clasificación y consistencia de la información, su aplicación supone que el desempeño y el comportamiento es común a los diferentes países. Si bien en términos generales esto podría ser cierto, los distintos marcos legales e institucionales indudablemente afectan el comportamiento individual de las personas, sus incentivos, etcétera.

El contexto social de cada situación, la cohesión social y otros factores similares no son fáciles de incorporar al marco de un modelo cuantitativo. Por ello, hay que tomar como punto de partida los parámetros generales, los cuales podemos modificar para reflejar las circunstancias locales. Esto refuerza la importancia de incluir a los actores locales cuando se aplican estas herramientas para que las variables de índole local sean tomados en cuenta. Modelos como el REMI y el LEFM permiten a los analistas de las evaluaciones y a los diseñadores de políticas a nivel local una mejor comprensión con base en los descubrimientos generales del análisis económico y de otras disciplinas.

Se discutió la complejidad inherente al uso de este tipo de modelos —que indudablemente requiere un mínimo nivel de experiencia—, ya que sus resultados pueden ser sensibles a cómo son aplicados y a las presuposiciones del caso. Como siempre, existe por supuesto el peligro del “*garbage in garbage out*”, o sea que entre basura, y por lo tanto, salga basura.

Otra preocupación fue la relativa a cuán oportunamente llegan los datos clave, especialmente algunos factores como por ejemplo las tablas de insumo-producto. Sin embargo, esta cuestión no es tan relevante, ya que la mayoría de estos valores evolucionan muy lentamente a través del tiempo. Por lo tanto, se insistió en la importancia de concentrarse en el análisis de los efectos relativos (comparar una política con otra, comparar su presencia con su ausencia) en vez de concentrarse en los cambios absolutos.

Utilidad de los modelos macro

Algunas intervenciones anteriores, particularmente la de Bartik, señalaron que cuando se evalúan políticas o programas particulares, es de suma importancia ver más allá de los impactos inmediatos a individuos o empresas. En particular, una evaluación completa debe analizar y valorar el efecto global de la política en los ingresos y niveles totales de empleo. Mientras que casi todos los modelos de evaluación defendidos en las intervenciones anteriores enfocan aspectos como el peso muerto o los desplazamientos a nivel micro, estos no ofrecen mecanismo alguno que permita estudiar los efectos a nivel macroeconómico. Igualmente, aunque los otros métodos proponen diversas formas de evaluar los impactos pasados de un programa, no logran vislumbrar sus posibles efectos a futuro.

A fin de llegar a una mejor comprensión, se necesita desarrollar un modelo de pronósticos macroeconómicos. Este modelo ofrecerá un mecanismo mediante el cual podremos llegar a las consecuencias macroeconómicas de los efectos individuales (vía la construcción de multiplicadores o efectos de vinculación insumo-producto), así como a consecuencias contrafactuales de no haberse implementado el programa en cuestión. La conclusión clave de la discusión fue que, si bien dichos modelos no son la panacea, pues sólo enfocan una parte y no la totalidad, pueden sin embargo ofrecernos una mejor comprensión del proceso de evaluación y nos hacen recordar las complejas interacciones que debemos tomar en cuenta pero que a veces olvidamos.

Algunas de las ponencias de la conferencia, por ejemplo la de Wong, enfatizaron la necesidad de tomar en cuenta los efectos del desplazamiento geográfico. Si bien para estos casos los modelos macro no funcionan por sí solos, proporcionan, sin embargo, un importante marco teórico.

Por otro lado, Bartik señaló que podemos hacer una evaluación rigurosa utilizando distintos métodos, siempre que estos vinculen en la econometría regional los impactos fiscales y el mercado laboral local. Se necesita una combinación de técnicas cuantitativas y cualitativas, con miras a mejorar y no a "aniquilar" los programas que son deficientes.

Asimismo, Smith señaló la importancia de la sinergia y de los efectos macroeconómicos, aspectos que pueden pasarse por alto en una evaluación netamente micro-económica.

Oldsman subrayó cuán importante es que los datos de referencia y la información inicial sean sólidos. Clave en este proceso es suponer qué hubiese sucedido de no haber modificado/implementado una política.

Eberts y O'Leary tocaron el tema de los programas sociales cuyo propósito es influir en la economía regional y local, y lo difícil que es evaluarlos. Sus alcances van más allá de un simple impacto en los actores individuales u organizados del

programa. Por otro lado, la escala de las intervenciones es a veces tan pequeña que sus efectos son imperceptibles. Un modelo macro es al menos un medio para poder abordar este tipo de resultados.

Walter destacó la necesidad de recolectar información longitudinal para facilitar la construcción de modelos y el desarrollo de mejores y más sofisticadas herramientas de pronóstico.

Por último, Stern demostró que las políticas y subsecuentes evaluaciones pueden interactuar y evolucionar, cambiando el comportamiento y generando capacidades adicionales. Lo anterior requiere un manejo cuidadoso del marco conceptual del modelo, ya que sus parámetros se basan en patrones de comportamiento previos y en marcos institucionales o relacionados.

Wong finalmente nos recordó que las intervenciones en las ciencias sociales no son simples repeticiones sistemáticas de experimentos controlados. A menudo, la realización del experimento, así como su evaluación, influirá en los resultados y en el comportamiento de los involucrados. Esto se aplica particularmente a los casos que utilizan modelos macro tipo REMI o LEFM. Los cambios de comportamiento podrían indicar la necesidad de revalorar los supuestos que sirvieron de base para fijar parámetros. Robson recalcó un punto similar, aludiendo a la importancia de una investigación orientada a las acciones y a los indicadores claros, con un enfoque cualitativo y flexible en vez de uno cuantitativo y rígido.

Referencia bibliográfica

WILSON, R., A. ASSEFA y J. BEARD (1995) "A Local Economy Forecasting Model (LEFM) for the UK Economy". Trabajo presentado en el Simposio Europeo sobre el Desarrollo de los Mercados Laborales, Universidad de Warwick, 18 y 19 de mayo de 1995.

Capítulo 8

Evaluación de las políticas de áreas

Por
Brian Robson,
Centro para los Estudios de Políticas,
Universidad de Manchester,
Reino Unido

El contexto de la política

Las primeras políticas urbanas formales en Gran Bretaña datan de finales de los sesenta y, durante las siguientes tres décadas, la política ha otorgado gran confianza en las Iniciativas de Áreas (Area-based initiatives, ABIs) como un mecanismo para enfrentar los problemas relacionados con el abandono y la marginación de los grandes pueblos y ciudades que fueron industriales en la Gran Bretaña (por ejemplo, Hall y Nevin, 1999).

Desde la década de los ochenta existe una colección desconcertante de diferentes tipos de instrumentos de política, entre los cuales han sido, o siguen siendo:

- El Movimiento en favor de las Ciudades (AfC), el cual enfocó recursos adicionales en 57 distritos de autoridad local (DAL) (LADs), en la década de los ochenta.
- Desafío de la Ciudad (CC), que operó entre 1992 y 1998, dirigió recursos a 31 áreas subdistritales.
- El Presupuesto Único para la Regeneración (SRB), que agrupó el gasto de 21 programas anteriores, sostuvo seis rondas de asignación durante la década de los noventa, y fue focalizado en gran parte a áreas pequeñas definidas geográficamente, las cuales recibieron recursos financieros hasta por más de siete años.
- Nuevo Trato para las Comunidades (NDC), que actualmente apoya a 39 áreas subdistritales con poblaciones de hasta 4000 habitantes durante un periodo de diez años.
- Empresas de Regeneración Urbana (URC), las primeras de las cuales fueron creadas en 1999, actualmente se cuenta con 11 establecidas para desarrollar los planes maestros, principalmente para las áreas del centro de las ciudades y con la ejecución de los gastos financiados a través de sociedades, entre Agencias de Desarrollo Regionales (RDA), Autoridades Municipales (LADs) y Sociedades Inglesas (EP).

Cada uno de estos programas ha sido financiado por el departamento gubernamental con la responsabilidad principal de la regeneración urbana, en lo

sucesivo llamadas Departamento del Medio ambiente (DoE), Departamento del Transporte Ambiental y de las Regiones (DETR), Departamento del Transporte, Gobierno Local y las Regiones (DTLR), y ahora la Oficina del viceprimer ministro (ODPM).

Adicionalmente, en los últimos cinco años, se ha registrado un número creciente de Iniciativas de Áreas Complementarias (ABIs) de otros departamentos del gobierno central. Entre los ejemplos se incluyen: el Programa de Reducción del Crimen, el de Inicio Seguro, Zonas de Acción Educativa, Zonas de Empleo, Zonas de Acción Deportiva, y Zonas de Acción de Salud, pertenecientes a los departamentos gubernamentales, tales como: el del Interior, el del Empleo, el de Educación, el del Deporte y Cultura y el de Salud.

Dichos enfoques de áreas están basados en el supuesto de que, por un lado, existen procesos relacionados con el área que agravan los problemas que enfrentan los individuos con carencias, o bien que existen eficiencias relacionadas con la provisión de la política en el interior de las áreas objetivo definidas, por el otro. Al mismo tiempo que se ha discutido mucho en la bibliografía académica sobre si un “efecto de área” puede demostrarse (Dorling et al., 2001; Atkinson y Kintrea, 2001), parece haber un acuerdo general sobre la existencia de beneficios administrativos asociados con el objetivo geográfico —sin dejar de considerar la sinergia potencial que puede alcanzarse a través de los diferentes ámbitos de política—, y la eficiencia de desplegar el personal y los recursos dentro de un número limitado de áreas. El caso del área objetivo es que, dadas las amplias disparidades en carencias, el vecindario o el barrio es la escala más apropiada para promover la identidad y la participación comunitarias, y que los recursos concentrados en pequeñas áreas durante cierto número de años pueden significar un paso en la mejora de las circunstancias de las áreas marginadas (Lawless et al., 2000).

El centro de atención de las Iniciativas de Áreas (ABIs) ha cambiado con el tiempo. Al principio, en los años setenta, la mayor parte de los programas cubrían objetivos económicos, sociales y ambientales en general, pero los recursos adicionales asignados eran generalmente pequeños. Posteriormente, en los años ochenta, la mayoría del financiamiento se dirigió al mejoramiento físico de las áreas abandonadas en un intento por reanimar el funcionamiento del mercado inmobiliario. Las intervenciones típicas incluyeron programas como el de Zonas Empresariales, en las que se ofrecían incentivos financieros a las empresas que decidieran ubicarse en las subáreas urbanas depreciadas, o como las Corporaciones de Desarrollo Urbano, administradas por consejos del sector privado destinadas a restaurar la infraestructura de los sitios abandonados que en algún momento fueron industriales. En la década de los noventa, el Desafío de la Ciudad y los programas SRB reintegraron a las autoridades locales y a su vez ampliaron el enfoque para abarcar aspectos sociales y económicos, así como objetivos de medioambiente y de la propiedad. Con este cambio de énfasis, dos características principales dominan los enfoques recientes de regeneración

urbana: el énfasis a trabajar en sociedad y, relacionado con esto, el objetivo de desarrollar una mejor coordinación entre los diversos ámbitos de la política.

El último Informe Oficial sobre desarrollo urbano (DETR, 2000) cambió fundamentalmente el empuje principal de la política. Se alejó de la anterior dependencia casi exclusiva sobre las iniciativas de áreas y se adhirió al objetivo de “la corriente principal” como una mejor forma para abordar los problemas urbanos. La corriente principal puede ser pensada de tres modos distintos:

- el intento de desviar los recursos de los principales programas de gasto (tales como educación, servicios sociales o vivienda) a las áreas objetivo con necesidades especiales o de mejorar la calidad con que se proveen los servicios a tales áreas;
- el esfuerzo por aprender las lecciones sobre lo que funciona en programas y proyectos específicos y su aplicación más general en otras áreas; y
- la tentativa de incorporar las enseñanzas de política que surgen de iniciativas específicas en los servicios principales.

Aun cuando la integración a la corriente principal ha sido por mucho tiempo un *mantra* dentro de la política urbana, el nuevo marco de trabajo de la política es el primero en poner aquélla en el centro del escenario. Las Sociedades Estratégicas Locales están ahora en proceso de convertirse en las autoridades locales y son las responsables del desarrollo de las Estrategias Comunitarias que utilizan meticulosamente la corriente principal, así como los recursos específicos provenientes de fondos como el Fondo de Renovación del Vecindario (dirigido a 88 autoridades locales) para revertir las condiciones de las áreas pobres.

Entonces, aun cuando muchas iniciativas de áreas se mantengan —no menos NDCs, URCs y algunas de las últimas rondas de financiamiento de SRB— la política urbana inglesa ahora ha dibujado una línea de dependencia casi exclusiva en las iniciativas de áreas.

El crecimiento del monitoreo y de la evaluación

Desde el principio de esta historia de política urbana de 30 años, ha existido un desarrollo continuo del monitoreo y de la evaluación bajo los auspicios del gobierno central. Los recursos totales canalizados en el monitoreo y la evaluación han crecido considerablemente. Ante esto, la evaluación actualmente se ha convertido en parte integral del ambiente de la política. Asimismo, los proyectos de evaluación auspiciados por los departamentos financieros del gobierno central, las revisiones genéricas de política, son emprendidas por estructuras tales como la Comisión Auditora (por ejemplo, Comisión Auditora, 1989; 2002), la Oficina Nacional de Auditoría; las recién establecidas unidades de coordinación del gobierno central,

como la Unidad de Exclusión Social y la Unidad de Coordinación Regional también usan la evidencia basada en investigación, con el objeto de conducir la dirección de la política. Un ejemplo reciente es la revisión de la Unidad de Coordinación Regional de iniciativas de áreas (RCU; 2002) que hace recomendaciones sobre la fusión e integración de varios programas de regeneración provenientes de una amplia gama de departamentos gubernamentales, que actualmente son financiados separadamente. Además de dicha evaluación patrocinada por el gobierno, han existido programas que evalúan las políticas de áreas realizadas por organizaciones externas de investigación —sobre todo el Programa de Regeneración de Áreas de la Fundación Joseph Rowntree (véase, por ejemplo, MacLennan, 2000)— y por el Consejo de Investigación Económica y Social a través de su Programa de Ciudades (véase, por ejemplo, Begg, 2002). En esta ponencia, sin embargo, la atención está restringida a la investigación auspiciada por el gobierno.

Gran parte de la investigación respaldada por el Departamento del Medioambiente en las décadas de los setenta y ochenta fue relativamente caótica. La mayoría de las evaluaciones fueron *ex post* y los investigadores padecieron ciertas dificultades para armar la información que pudiese, reconocer las condiciones iniciales y recrear estos datos en los resultados durante la vigencia de las iniciativas. Sin embargo, desde la década de los noventa (y especialmente desde los programas de Desafío de la Ciudad y los de SRB) en general, la mayor parte del financiamiento de las intervenciones de las políticas ha dependido de la supervisión y evaluación continuas de las propias asociaciones locales, lo cual ha sido acompañado por evaluaciones nacionales realizadas para la ODPM. El intenso cambio en la evaluación se dio con la llegada del nuevo gobierno en 1997. Ahora existe una cultura de la evaluación local y nacional firmemente arraigada en casi todas las Iniciativas de Áreas.

Este aumento en la evaluación es un reflejo del énfasis creciente del gobierno por buscar “valor por dinero” y sobre su *mantra* de “lo que importa es lo que funciona”. La noción de la creación de políticas basada en pruebas se ha visto de manera más clara en el campo de la medicina (donde, por ejemplo, la Colaboración Cochrane y unidades como el Centro para la Medicina Basada en Pruebas en Oxford han desarrollado una gama de análisis de estudios basados en pruebas sobre la atención a la salud). Convertir tal experiencia en la evaluación de los impactos de la regeneración basada en áreas ha representado desafíos muy importantes a las ciencias sociales.

La evaluación de 1970 a 1990

De una manera un tanto insolente, la urgencia oficial de monitorear y evaluar se convirtió rápidamente en institucional y rutinaria perdió así su vanguardia. Por ejemplo, Ho ha indicado una tipología compuesta de tres partes para el periodo entre la década de los setenta y los noventa, lo que ella llama las tres eras de

la investigación oficial de evaluación: “Inocencia” a comienzos de la década de los setenta; el “Disentimiento” en los ochenta; y el “Consentimiento” durante los noventa (Ho, 1999). Más que cerrar el ciclo entre monitoreo/evaluación/reformulación de políticas, argumenta que, con el tiempo, la evaluación fue usada cada vez más para justificar y aplaudir lo que ya se había hecho.

Tres proyectos de evaluación sucesivos pueden ser usados para ilustrar este cambio en el balance de los tipos del enfoque.

Proyecto de Desarrollo Comunitario

El primero de los proyectos es el más ilustrado por las aleccionadoras evaluaciones del Proyecto de Desarrollo de la Comunidad (CDP), un programa que registró un fuerte énfasis sobre la investigación de acción cuyo objetivo declarado era influir en la formulación de políticas. Los doce proyectos CDP establecidos por el Ministerio del Interior en los años setenta fueron un experimento importantísimo para mejorar los servicios sociales para aquellos más necesitados. Se emplearon equipos de acción por las autoridades locales respectivas con financiamiento del Ministerio del Interior, al igual que la Unidad de Información e Inteligencia Central (Higgins et al., 1983). Cada equipo produjo estudios exhaustivos de sus áreas del proyecto y una serie de informes inter-proyecto fue publicada, la cual proporcionó diagnósticos de las causas de la pobreza y la exclusión. Dichos diagnósticos fueron muy críticos del enfoque tan estrecho y a pequeña escala de la política gubernamental de esa época, sin desestimar que estaba basado en una filosofía de patología social que suponía que los problemas eran internos en las comunidades pequeñas más que estar arraigados en el funcionamiento del amplio contexto político y económico (CDP, 1977). Considerando la dirección tomada por los equipos de investigación y la naturaleza crítica de sus informes, quizás no fue ninguna sorpresa que la Unidad de Información e Inteligencia central fuese cerrada prematuramente en 1976 y que el programa de CDP fuese clausurado.

Acción para las Ciudades

Un segundo ejemplo es la evaluación de DoE sobre la recolección de iniciativas incorporadas bajo el encabezado de Acción para las Ciudades (AfC) a finales de los ochenta (Robson et al., 1994). Este proyecto intentó desarrollar una metodología para abordar algunos de los principales acertijos de la evaluación cuantitativa que fueron resumidos como las seis Cs:

Aspectos **contrafactuales** (en la situación contraria), o de pérdida social. Lo que habría ocurrido de todos modos, incluso en la ausencia de la intervención del gobierno.

Efectos de **contigüidad** o de desplazamiento. Los impactos de las políticas en las áreas objetivo podrían generar efectos positivos o negativos en áreas adyacentes.

Aspectos de **confusión**. Los resultados pueden ser ocasionados por diversas iniciativas, a menudo traslapadas, de forma tal que es difícil atribuir de forma exclusiva el cambio total a alguno de los múltiples programas.

Efectos **contextuales**. Los programas inician desde niveles muy diferentes de activos y de capacidades; así, los cambios nacionales más amplios pueden tener impacto en las localidades de maneras muy diferentes.

Aspectos **combinatorios**. Los paquetes de intervención incluyen una variedad de diferentes combinaciones de programas —dirigidos a la creación de empleos, el mejoramiento físico, la reducción del crimen, la mejora de salud y otros semejantes— cada uno de los cuales puede tener efectos desencadenantes sobre los otros. Algunas combinaciones demuestran ser más eficaces que otras.

Efectos de **elección**. El conjunto de los lugares elegidos para programas específicos cambian con el tiempo y a través de diferentes programas, y por ello se torna difícil asignar inequívocamente políticas inactivas o políticas activas a lugares específicos.

La necesidad de tomar en cuenta estos aspectos fue subsecuentemente incorporada en las pautas formales del erario público para la investigación de la evaluación (HM Treasury, 1995). La Coordinación del Erario Público ha “incentivado a las agencias y a las sociedades a ser más concisas en términos de la definición de los objetivos e insumos, a ser más sensibles ante la importancia de establecer resultados netos, en vez de los brutos, y a equiparse mejor para evaluar el valor por dinero” (Lawless et al., 2000). Sin embargo, mientras la Coordinación parece suponer que estos asuntos ya pueden ser abordados fácilmente por la investigación evaluativa, en la práctica dichos asuntos continúan representando verdaderos acertijos para las metodologías de evaluación, por lo que gran parte de la evaluación todavía se concentra en la producción en lugar de los resultados y tiene grandes problemas para observar la interrelación entre las diferentes ramas de la política.

Dado que la investigación del AfC fue probablemente la evaluación general más ambiciosa en las décadas de los ochenta y noventa, vale la pena mirar en detalle su enfoque. Dicho enfoque utilizó métodos tanto cuantitativos como cualitativos. El aspecto cuantitativo observó la relación entre los insumos financieros y los resultados socioeconómicos usando el análisis de regresión y medidas de concentración espacial. Los insumos comprendían tanto los recursos dirigidos al programa urbano como también los recursos principales canalizados hacia las autoridades locales. Los resultados se midieron mediante cinco

indicadores de alto nivel: el desempleo, la generación de empleos, la creación de pequeñas empresas, el cambio de precios de la vivienda y la migración neta de individuos entre 25 y 34 años. A través de la observación de la relación entre los insumos y estos indicadores de resultados de alto nivel, la investigación llegó a dos supuestos tácitos: que era imposible establecer un grupo de comparaciones entre “políticas inactivas” y “políticas activas” —no sólo porque la política cambia constantemente, sino también porque casi todos los lugares marginados habían recibido una u otra forma de intervención de políticas; y que en la compleja arena política de la regeneración era difícil aislar los impactos de las intervenciones específicas de las políticas. Por lo tanto, el enfoque de utilizar un diseño cuasi-experimental supuso que más recursos implicaban más intervención política y la hipótesis nula era que no existiría ninguna relación entre los insumos y los resultados.

El análisis fue llevado a cabo en una variedad de escalas espaciales. En una escala de autoridad local, se observaron los resultados de 123 LADs: las 57 áreas urbanas prioritarias (UPAs) que fueron receptoras del financiamiento directo dirigido; 40 áreas “marginales” con condiciones no distintas de las UPAs; y 26 municipios de “comparación” los cuales no recibieron ningún recurso adicional. La relación entre la aplicación de recursos y los cambios en los indicadores de resultados socioeconómicos fue utilizada para generar una medida global del impacto total de la política, con el objetivo de probar si el tener más recursos se relacionaba con las mejoras absolutas o relativas en las condiciones de las áreas.

En una escala amplia se vieron cambios en la disparidad entre subáreas pobres y no marginadas en tres conurbaciones; se intentó también —de forma infructuosa— utilizar la modelación multinivel para desenmarañar las relaciones a diversas escalas espaciales.

Además, el análisis también incluyó una gama de encuestas y entrevistas a las tres conurbaciones seleccionadas para observar los procesos subyacentes de los intentos en la regeneración. Éstos implicaban un cuestionario a gran escala de residentes y entrevistas con empleadores y “expertos” en políticas.

Las conclusiones sugerían que la política había tenido resultados combinados. Por un lado, para la mayor parte de los indicadores de resultados hubo poca mejoría relativa en las áreas objetivo de las políticas. Las condiciones en las peores áreas (especialmente los núcleos de las conurbaciones) se deterioraron y la creciente concentración espacial de la pobreza y del desempleo indicó que el nivel de la exclusión social aumentó. Por otro lado, los distritos objetivo mostraron la mejora relativa en el empleo y en el éxito de atraer inmigración neta de trabajadores jóvenes, los residentes eran más optimistas respecto de las perspectivas de sus áreas donde operaban las intervenciones de ABI, y también había una relación positiva entre las cantidades de los recursos dirigidos y la mejora relativa de

las áreas. En conjunto, las intervenciones basadas en áreas parecieron ser beneficiosas, pero no para hacer mejoras importantes en las peores áreas.

Esta evaluación cae en la categorización del “Disentimiento” de Ho. En dicha categoría se recomendaba una variedad de cambios a la política gubernamental —una mejor coordinación entre los programas, principios más claros en la asignación de los recursos, el establecimiento de un “recipiente” urbano de recursos que pudieran ser usados más flexiblemente de acuerdo con las circunstancias locales, más incentivos para trabajar en sociedad entre las agencias. Muchas de estas recomendaciones fueron incluidas subsiguientemente en los principios del Presupuesto Único para la Regeneración y en este grado, puede argumentarse que la evaluación fue formativa así como también acumulativa, independientemente de la intención inicial del gobierno.

Desafío de la ciudad

El tercer ejemplo es la evaluación del Desafío de la Ciudad que apareció como un informe intermedio (Rusell, 1996) y un informe final (KPMG, 1999). Ambos reportes se concentraron esencialmente en los procesos en lugar de los resultados y en los procesos involucrados en la iniciativa. Se analizaron los datos de los resultados en las 31 áreas del Desafío de la Ciudad y se llevó a cabo una amplia gama de entrevistas. Al igual que las evaluaciones acumulativas, los cálculos de las evaluaciones citadas estaban basados en los procesos y las estructuras, en lugar de los resultados, por lo que las conclusiones tienen el tinte de alabanza de los logros de un programa que fue ampliamente pregonado como un valioso cambio respecto del enfoque más estrecho de la década de los ochenta sobre regeneración física y marginación de las autoridades locales.

El informe intermedio, por ejemplo, desarrolló cinco líneas principales:

- la recolección de los datos iniciales;
- gasto y procesos de las 31 áreas del Desafío de la Ciudad, analizando la descomposición del gasto, las fuentes de financiamiento y los procesos en relación con las metas anuales;
- estudios de caso más detallados en 14 de las áreas, recurriendo a las entrevistas con funcionarios y partes interesadas de los sectores público, privado y comunitario, junto con información documental;
- un cuestionario por correo postal dirigido a socios clave de las 31 áreas, que incluía preguntas sobre desplazamientos así como sobre las estructuras y los procesos del programa; y
- estudios de caso de dos áreas que habían licitado para ganar el estatus del Desafío de la Ciudad, pero fallaron en obtenerlo.

La tendencia final fue el elemento más novedoso del proyecto. Al revisar lo que había o no había ocurrido en las áreas que fallaron en conseguir los recursos del Desafío de la Ciudad, la evaluación fue capaz de ilustrar las inferencias sobre la adicionalidad asociada al programa. Por ejemplo, para una de las autoridades no exitosas, la falla al obtener los recursos del Desafío de la Ciudad significó que otros compromisos para conseguir la licitación se perdieran, por lo que otras áreas marginadas fuera de la zona del Desafío de la Ciudad tuvieron que ser movidas aún más abajo en la lista de prioridades de la autoridad local, por lo que existía una brecha fundamental entre la ejecución de un sistema de proyectos (algunos financiados a través de otros recursos dirigidos) y la sinergia que habría sido conseguida con un programa acordado a largo plazo como el Desafío de la Ciudad. Cada uno de estos puntos proporciona algunas pruebas sobre la adicionalidad asociada con el programa de Desafío de la Ciudad.

Lo arriba expuesto y la evaluación más directa de las áreas financiadas, llevaron al equipo a argumentar que el Desafío de la Ciudad brindó una variedad de beneficios: el incentivo de desarrollar más planificación estratégica; el valor de desarrollar proyectos emblemáticos; el logro de una masa crítica de actividad de la cual fluyeran beneficios vinculados; y la sinergia que podía ser creada a través de los diversos campos políticos.

Mientras gran parte de los informes intermedios y finales fueron expresados en términos de enlistar los logros del Desafío de la Ciudad —los niveles de apalancamiento financiero alcanzados, el número de viviendas construidas o mejoradas, la conservación o creación de empleo, el reclamo o mejora de las tierras abandonadas, espacio creado o mejorado para industrias y oficinas, la promoción de la generación de negocios— también tuvieron mucho que decir sobre los procesos y las estructuras mediante las cuales el programa fue proveído. Sería un juicio demasiado severo argumentar que tales evaluaciones han sido simplemente “condescendientes”. En la medida en que la política ha abordado temas interconectados más amplios, quizás se ha vuelto inevitable darle un énfasis más fuerte a la evaluación cualitativa. En consecuencia, hemos aprendido mucho sobre el papel de las estructuras, el funcionamiento de sociedades y la trascendencia de los individuos.

Existe poca duda de que, a partir de la ahora larga historia de monitoreo y evaluación, las sucesivas administraciones de los gobiernos han aprendido mucho sobre los desafíos de la regeneración y sobre las formas para desarrollar mejores enfoques que enfrenten a los elementos entrelazados de la decadencia y la marginación urbanas. Los cambios políticos sucesivos han reflejado este camino por la curva de aprendizaje, al menos en la constitución de un enfoque más centrado en la comunidad para la regeneración, el énfasis creciente en la coordinación entre los diferentes ámbitos políticos y el énfasis en las sociedades trabajando en el territorio.

Evaluaciones recientes y actuales

Desde 1997 ha existido un verdadero auge en evaluación. Esto se ve reflejado en la escala de recursos ahora canalizados hacia la investigación de evaluación, con más de ocho millones de libras destinados a la evaluación de los programas de regeneración por la ODPM en el ejercicio actual (Tabla 8.1). Esto se atribuye en gran medida a la naturaleza pragmática no ideológica de la nueva administración del gobierno. Su enfoque de “lo que funciona” y su creciente énfasis sobre la dimensión del cumplimiento de la política han ayudado a impulsar una combinación de evaluaciones formativas y acumulativas. Las evaluaciones nacionales de los programas generales, a la par del requerimiento de que las asociaciones locales de regeneración dirijan su propio monitoreo y evaluación, han generado estudios en abundancia. Esto ha sido relacionado con un enfoque que ha introducido muchas iniciativas políticas a través de programas piloto en áreas seleccionadas que subsecuentemente puedan ser dirigidas a niveles más generales a otras localidades. Además, el gobierno ha brindado un acceso mucho más amplio a la información (a través de la aparición de la Base de Datos Nacional de Estadísticas del Vecindario) y a los resultados de los proyectos de evaluación, la mayoría de los cuales han sido colocados en los sitios web del gobierno.

Tabla 8.1 Gasto de la Oficina del viceprimer ministro en programas de regeneración e investigación de evaluación, 2002

	Libras Esterlinas
Programas sustantivos de regeneración	2 366 000 000
Investigación de regeneración	10 000 000
(de los cuales) investigación de evaluación	8 700 000
(de los cuales) evaluación del Nuevo Trato a Comunidades	5 550 000

Fuente: comunicación privada, Oficina del viceprimer ministro (ODPM).

El trabajo que ha surgido de esto encarna una amplia gama de tipos de evaluación: descriptiva, analítica, teórica, perceptiva y diagnóstica. Lo que ha sido especialmente evidente es el cambio en la manera en la que la relación entre la investigación y la práctica ha ido evolucionando. Más que observar un solo continuo desde la investigación pura hasta el desarrollo de la estrategia para los cambios en la práctica, ahora existe una mayor comprensión de la pluralidad en la relación entre la evaluación y la práctica. Por lo tanto ahora se presenta una combinación más grande de la investigación de acción y de evaluación con equipos que trabajan más de cerca con los operadores de la política en el campo. Algunas de las principales evaluaciones vigentes incluyen:

- La evaluación del programa SRB (por ejemplo, Rhodes *et al.*, 2002). Éste es un proyecto a largo plazo en el que se han desarrollado análisis cuantitativos de

los procesos y los resultados, observando a los proyectos específicos que son parte de los programas de sociedades locales e investigando la eficacia de las sociedades mismas.

- La coordinación de iniciativas en áreas con múltiples intervenciones políticas (Stewart *et al.*, 2002). Esto se centra en seis áreas que han sido receptoras de casi todos las ABIs y trabajando con las autoridades relevantes, en esencia toma la forma de investigación de acción, además de que ofrece una visión acumulativa de los resultados de los programas.
- El proyecto de 24 ciudades (Falk, 2002). Este es un estudio cualitativo de 24 distritos seleccionados para desarrollar críticas de las "visiones" de las sociedades locales y para identificar ejemplos de las mejores prácticas que puedan ser transferibles a otro lugar. Para ello se han realizado visitas, talleres con los ciudadanos, los jóvenes y grupos de propiedades para identificar las prácticas ejemplares. Se informó sobre sus hallazgos iniciales a la Cumbre Urbana que realizó el gobierno en octubre 2002.

Nuevo Trato a las Comunidades (NDC)

La evaluación nacional del Nuevo Trato a las Comunidades es la más ambiciosa de las iniciativas en curso, la cual se lleva a cabo bajo los auspicios de la Unidad de Renovación del Vecindario en la ODPM. El financiamiento para este trabajo fue incluido desde el principio en el presupuesto nacional del NDC y una serie de equipos de investigación está siendo coordinada a través de la Universidad de Sheffield Hallam. Este importante proyecto combina elementos de la evaluación formativa, acumulativa y de ajuste (utilizando las categorías sugeridas por Rossi y Freeman, 1999). Están involucrados tres tipos de equipos de evaluación:

- equipos de las asociaciones, quienes observan el trabajo de cada una de las sociedades individuales del NDC;
- equipos temáticos de corte transversal quienes revisan los distintos campos políticos a través de las 39 áreas del NDC; y
- equipos complementarios que examinan los aspectos técnicos del programa.

Cada uno de los equipos de las asociaciones dirige un análisis de los objetivos y los logros de las asociaciones del NDC. Esto implica una crítica de los planes de entrega, de los proyectos seleccionados, de las estructuras establecidas en cada área, del grado en que se dan la integración de la corriente principal, y de la participación de la comunidad en el proceso de regeneración. Debido a que una de las características del NDC es su intento por incorporar a las comunidades más firmemente en la determinación de las prioridades y en el cumplimiento de la regeneración, el éxito de incentivar la participación de los grupos residentes

y de escuchar la voz de las comunidades son características centrales de las evaluaciones. Por ello, los equipos han usado una variedad de entrevistas con los que hacen las políticas, con los que la administran y con las partes interesadas relevantes, junto con grupos focales extraídos de los grupos de residentes y de las comunidades. Cada uno de los equipos produce las evaluaciones en un formato común con plantillas diseñadas por el centro, así como también interpretaciones más individuales del progreso en el territorio de las áreas individuales del NDC. El estilo del trabajo es en parte acumulativo y formativo, éste último ofreciendo consejos y retroalimentación a las asociaciones y a las ODPM como el departamento gubernamental patrocinador.

Los equipos temáticos cubren temas transversales como la falta de empleo, la salud, la delincuencia y la educación. Los equipos están desarrollando evaluaciones a través del análisis de datos primarios de las 39 áreas del NDC y de datos secundarios sobre los resultados, así como también están seleccionando proyectos específicos de muestras de NDCs para identificar ejemplos de las mejores prácticas que puedan ser transferibles a otros lugares.

Los equipos complementarios están conformados por tres grupos: uno desarrolla la evaluación "tradicional" de valor por dinero en términos de gastos y resultados; un segundo equipo junta los datos secundarios (especialmente sobre pagos de beneficios de la asistencia social) para residentes dentro de cada uno de los NDCs; y un tercer equipo utilizará los resultados del próximo censo de 2001 para desarrollar puntos de partida formales de los indicadores al inicio del programa NDC, con vista hacia las comparaciones eventuales que pudiesen realizarse al final del programa NDC. Además una organización comercial de encuestas ha conducido una gran encuesta en los hogares dentro de cada una de las 39 áreas para producir datos sobre las actitudes y las circunstancias de los residentes. Una o más encuestas de seguimiento serán emprendidas en etapas posteriores durante la vigencia del programa general.

En esta fase, la evaluación nacional está programada para continuar durante 2005 y existe la expectativa de que pueda cubrir los 10 años de vigencia del programa NDC como un todo.

Esta evaluación nacional está siendo dirigida a la par de las evaluaciones locales que cada una de sociedades del NDC está obligada a llevar a cabo como parte de su trabajo. Dichas evaluaciones locales no son realizadas bajo un formato común, sino que implican una combinación de encuestas en los hogares, y los análisis de los procesos y de los resultados. En comparación con la evaluación nacional, las evaluaciones locales ponen un énfasis mayor en la evaluación de proyectos individuales y se centran menos en los resultados.

Con el tiempo, se ha notado un aumento dramático en la actividad evaluativa en Gran Bretaña que ha cambiado en forma y fondo. Mientras que se ha aprendido

mucho sobre las metodologías relevantes, se ha dado un distanciamiento de las evaluaciones cuantitativas descendentes más formales y se orienta más hacia los enfoques ascendentes cualitativos que han dependido de las entrevistas con los administradores y los receptores de la política, de los grupos focales y paneles, y de encuestas sociales.

Datos geográficos y focalización

Está claro que uno de los desafíos prácticos más importantes a los que se enfrentan todas estas evaluaciones ha sido la necesidad de acceder a datos apropiados en escalas espaciales relevantes. Debido a que muchas intervenciones políticas son dirigidas a áreas relativamente pequeñas, la evaluación eficaz de los resultados requiere de datos de áreas pequeñas que puedan ser comparados con el tiempo. Sin embargo, hasta hace poco (y con excepción del censo de la década), ha habido una falta de dichos datos sobre los cuales se pudiese realizar una investigación. La mayor parte de los datos administrativos pertenecientes a los departamentos del gobierno central se han aplicado a las autoridades locales —una escala demasiado tosca para interpretar los impactos de la mayoría de las iniciativas basadas en áreas. Muchos de los tipos de información relevantes (por ejemplo sobre vivienda, los aspectos de la fuerza de trabajo, o la percepción sobre el crimen) sólo se pueden obtener de encuestas nacionales cuyo tamaño de la muestra ha sido demasiado pequeño para producir datos contundentes, aun en una escala distrital, sin mencionar una subdistrital. Gran parte de la información de procesos proviene de los datos administrativos locales recabados por las propias agencias de regeneración y esto tiene la desventaja de que no se da necesariamente una correspondencia entre las categorías utilizadas por aquellos que arman los datos o a través de los métodos de recolección de datos usados por distintas agencias.

Esta situación está cambiando ahora. El gobierno ha establecido una Base de Datos Nacional de Estadísticas del Vecindario (neighbourhood.statistics.gov.uk), la cual usa a los distritos de las autoridades locales como su estructura básica para reportar datos, y está comenzando a armar la información disponible al público a lo largo de diferentes conjuntos relevantes de datos. Hoy en día, también se está armando una serie creciente de datos administrativos sobre la base de códigos postales que proporcionen una configuración relativamente flexible mediante la cual agregar información a una variedad de marcos de trabajo de áreas pequeñas. Por ejemplo, para datos sobre el rendimiento educativo de alumnos, antes de desarrollar los datos de códigos postales y los resultados de los exámenes escolares habían estado disponibles solamente escuela por escuela. Esto cubrió la secuencia de pruebas de “escenario clave” aplicadas a los niveles de la escuela primaria y secundaria, así como los exámenes formales aplicados a los 16 años de edad (los exámenes de GCSE) y a los 18 años (exámenes nivel-A).

Para observar el rendimiento de alumnos en áreas específicas, los investigadores se enfrentaron con dos alternativas. Los investigadores podían asumir por una parte que los alumnos asistían a la escuela más cercana a ellos y así podían usar los datos basados en escuelas para atribuir los logros por áreas sobre un principio del vecino más cercano: una suposición un tanto heroica, aunque más cierta para los alumnos de primaria que para alumnos de secundaria. Alternativamente, los investigadores podían modelar los datos basados en escuelas girando en torno a la información de captación para obtener muestras de escuelas y así calcular el rendimiento de los alumnos, basado en áreas. Un ejemplo de esto último es el Índice de Carencias Múltiples (Noble, 2000). Ahora, sin embargo, el Departamento para la Educación y las Destrezas está comenzando a integrar los datos basados en la dirección del alumno y, una vez que la secuencia de datos exista, ésta permitirá una base más realista sobre la cual evaluar los cambios en los logros educativos de alumnos.

Una segunda dificultad ha sido que algunos departamentos gubernamentales usan distintas zonas geográficas administrativas de las cuales reportan sus datos. Las estadísticas de salud, por ejemplo, tienden a referirse a las áreas de salud que no coinciden con los distritos de las autoridades locales. De forma semejante, los datos delincuenciales reportados están disponibles, generalmente, para los módulos y distritos de la policía, mismos que coinciden sino imperfectamente con los distritos y las demarcaciones de las autoridades locales. Se ha hecho un uso difundido de los sistemas de información geográfica (GIS) para resolver tales conflictos de zonas geográficas, pero las estimaciones carecen de la solidez de los datos recolectados específicamente para áreas espaciales comunes.

El movimiento para mejorar la disponibilidad de los datos de áreas pequeñas ha surgido en parte por el énfasis del gobierno sobre una escala de vecindario en diversas iniciativas políticas. La Estrategia Nacional para la Renovación del Vecindario (Exclusión Social Unit, 2001) ha puesto en marcha una gama de intervenciones basadas en el vecindario para mejorar las perspectivas y la administración de los vecindarios. Esto ha ayudado a centrar la atención en la necesidad de datos a escala subdistrital. Eso también ha sido estimulado por el énfasis en dirigir espacialmente los recursos de regeneración, donde el gobierno ha hecho uso de una secuencia de índices de necesidades. Los primeros dos índices de necesidades (Robson, 1995; 1998) fueron desarrollados esencialmente a escala distrital, pero incorporaban datos de barrios (con poblaciones generalmente menores a los 10 000 habitantes) y poblaciones menores (generalmente en los cientos de habitantes) (Distritos de Enumeración – ED), del censo de 1991. Se incrementó un número de innovaciones: la definición de carencias múltiples como un conjunto de una pequeña cantidad de “campos”, cada cual medido por una gama de indicadores; el uso de valores Chi-cuadrado como medios para estandarizar los resultados de los eventos entre los diferentes indicadores; la

producción de resultados sobre las carencias para escalas espaciales agrupadas (de ED a barrio y a distrito); y una gama de medidas basadas en el grado (la sumatoria de valores) en la extensión (la proporción de un área con resultados por encima de un valor límite) y en la intensidad (el valor promedio de los tres peores barrios en un distrito).

El Índice de Carencias Múltiples 2000 —el índice actual del gobierno— fue desarrollado por un equipo en Oxford. Fue construido sobre estas innovaciones, lo cual fue muy importante por la posibilidad de integrar valores actualizados a una escala de barrio, e igualmente importante por la disponibilidad de datos de los barrios en temas como los beneficios de la asistencia social (Noble, 2000). Se utilizan seis “campos” de carencias —ingreso, empleo, salud, educación, vivienda y acceso a servicios. La metodología utilizada para calcular los resultados para los barrios brinda una reflexión de la disponibilidad de los datos en una escala de barrios:

Las medidas de **ingreso** y de **empleo** podían ser calculadas por el equipo de Oxford recurriendo a datos directos sobre el número de demandantes a lo largo de la gama de beneficios de la asistencia social, con el objeto de producir la proporción de la población total que calificaba para uno u otro beneficio en función de la necesidad.

La salud incluye datos basados en los barrios de los receptores de beneficios por incapacidad y una medida de tasas estandarizadas de mortalidad a nivel distrital.

La educación incluía un procedimiento complejo de modelo para calcular el rendimiento de los alumnos con base en los datos de escuelas.

La vivienda podía ser medida únicamente a nivel de barrio por un cálculo de las viviendas inadecuadas basándose en una muestra nacional relativamente pequeña de la Encuesta de las Condiciones de los Hogares Ingleses, junto con los datos del censo de 1991.

El acceso a servicios (un campo cuyos valores probaron relacionarse negativamente con el resto de los campos) utilizó una distancia en línea recta para una gama de servicios (escuelas primarias, médicos, oficinas de correo y tiendas de comida) calculada para los receptores de los beneficios de asistencia social.

Los conjuntos de indicadores para cada uno de los últimos cuatro campos fueron combinados sobre la base de ponderaciones obtenidas de análisis factoriales; las medidas de los campos fueron estandarizadas clasificando los valores y usando una transformación exponencial; la medida global fue calculada sumando los valores de los seis campos utilizando ponderadores predeterminados de 25, 25, 15, 15, 10 y 10.

Mientras que el Índice es el más detallado que se ha producido hasta la fecha, está claro que las limitaciones de los datos (así como las dificultades metodológicas asociadas con la producción de los resultados de los eventos compuestos), exceptuando los datos de los beneficios, todavía hacen problemáticos este tipo de índices. Por ejemplo, la falta de datos de delincuencia y de medidas de abandono físico son brechas reconocidas por la ODPM.

Estos índices han sido ampliamente utilizados por el gobierno para ayudar a dirigir los recursos a áreas consideradas como necesitadas. Por ejemplo, los primeros índices ayudaron en la selección de proyectos de SRB y el IMD (Índice de Necesidades Múltiples por sus siglas en inglés), guió la distribución de los Fondos de Renovación del Vecindario.

Los índices también han sido utilizados para evaluar la eficacia de la focalización espacial (aunque desde el momento que son unos de los determinantes de la focalización, se ha dado cierto grado de circularidad). La evidencia de los patrones de gasto de las tres primeras rondas del Presupuesto Único para la Regeneración indicaban que existe una poderosa relación positiva entre el gasto y la marginación socioeconómica (Tyler et al., 1998). De las entonces 366 autoridades locales, el 30 % del financiamiento se destinó a los 20 más marginados, el 63% a los 56 de mayor marginación y el 81% a los 99 más marginados. Las 267 autoridades restantes recibieron solamente 19% del financiamiento total —generalmente para pequeños gastos de carencias dentro de distritos relativamente prósperos en tres áreas.

Un ejemplo adicional del intento por desarrollar mejor información de áreas pequeñas es el trabajo de compilar los datos sobre la incidencia espacial del gasto dentro de las autoridades locales (Bramley et al., 1998). Esta investigación utilizó tres áreas para estudios de caso —Brent, Liverpool y Nottingham— para analizar el gasto público localmente relevante (seguro social, salud, educación, vivienda, transporte, seguridad pública y otros servicios del gobierno local) a escala de barrio, cubriendo de esta forma aproximadamente el 70% del gasto público total. Para esto se acudió a datos de código postal de los registros administrativos, información de las encuestas sobre los traslados a las instalaciones públicas, encuestas de los hogares para calcular tasas de utilización, la distribución del gasto en el GIS y las ubicaciones geográficas individuales para algunos planes de inversión mayores. Lo anterior permitió demostrar que hay una amplia variación en el gasto entre los barrios individuales, que el gasto en los barrios más carentes está aproximadamente 45% por arriba de los barrios menos carentes, pero que hay diferencias significativas entre los departamentos gubernamentales en su configuración del gasto en relación con la carencia. La habilidad para intercalar tales datos de gastos es evidentemente crítica para muchas evaluaciones de las iniciativas basadas en áreas. El hecho de que es posible —aunque enormemente demandante en tiempo y costos— es al menos alentador para la futura evaluación

del impacto de las políticas de regeneración basadas en áreas y para la evaluación del éxito de la “corriente principal”.

Problemas conceptuales

Varios problemas han persistido en plantear ciertas necesidades especiales en este conjunto tan elaborado de evaluaciones. Uno de ellos es la dificultad para desarrollar análisis longitudinales genuinos. Casi todas las evaluaciones de áreas son esencialmente comparaciones de corte transversal de las áreas en dos o más momentos en el tiempo. Esto, por supuesto, hace caso omiso del hecho de que parte de cualquier cambio socioeconómico en las áreas es generalmente el resultado de la movilidad de las familias entre las dos fechas, en lugar de ser el resultado de los cambios en las circunstancias de los residentes iniciales. Éste es un problema específico, ya que la regeneración frecuentemente incentiva, o efectivamente se propone, atraer a nuevos residentes. Hay ejemplos innumerables donde la creación de nuevos trabajos tomados por los residentes locales los motiva a mudarse a otro lugar de manera que, por ejemplo, los niveles de desempleo en la zona pudiesen mantenerse, aunque algunos de los residentes anteriormente desempleados estén ahora empleados. De igual forma, las mejoras físicas para un área (no menos que la construcción de viviendas nuevas) frecuentemente atraen a residentes nuevos que frecuentemente son más prósperos, con un mayor nivel educativo y con mayores posibilidades a estar empleados que las familias autóctonas; por lo tanto se eleva el nivel de los indicadores socioeconómicos basados en áreas aunque esto pueda no reflejar ninguna mejoría en las circunstancias de los residentes autóctonos. Idealmente, si el centro de interés está en los residentes en vez de las áreas *per se*, las evaluaciones debieran rastrear los cambios a los residentes iniciales mediante algún tipo de encuesta longitudinal de individuos. Por ejemplo, la continua evaluación del programa del Presupuesto Único para la Regeneración ha usado encuestas sociales múltiples como un enfoque valioso para medir el cambio (Rhodes *et al.*, 2002). Claramente, esto significa considerables problemas logísticos, especialmente el desafío de rastrear a aquellos que se han mudado fuera de las áreas objetivo. Se han hecho esfuerzos para atacar el problema usando amigos y vecinos como fuentes de información sobre el paradero de aquellos que se han mudado y a través del uso de paneles continuos de información sobre los residentes (por ejemplo en las evaluaciones de SRB de Rhodes *et al.*, 2002), pero la tasa de éxito en el rastreo de los individuos que se mudan de las áreas ha sido comprensiblemente limitada. Las mismas dificultades han sido experimentadas con los sucesivos grupos de panel que han intentado mantener involucradas a las mismas personas en rondas sucesivas.

Irónicamente, en Gran Bretaña abundan las encuestas longitudinales —los ejemplos incluyen el Estudio Nacional del Desarrollo del Niño y los Datos Longitudinales que han sido recolectados para muestras de individuos idénticos

en los censos de población sucesivos desde 1981. McCulloch (2001), por ejemplo, ha manejado el Estudio de Panel de los Hogares Británicos para explorar el papel de las características individuales *versus* características basadas en áreas como enfoque para enfrentar la marginación. Sin embargo, para las evaluaciones basadas en áreas dichas muestras nacionales son demasiado pequeñas para ser usadas en el análisis de los cambios a escala de barrios o de vecindarios específicos en los que se aplican los programas de regeneración.

Un segundo problema es el cálculo preciso de la pérdida social y del desplazamiento. La mayoría de los estudios de pérdida social han empleado cálculos derivados de sus autoevaluaciones como medida, preguntando a empresas u otras agencias cuán lejos se habrían dado los procesos en ausencia de la intervención de las políticas (por ejemplo, Lenihan, 2001). Evidentemente hay cuestiones de interés propio involucradas en las respuestas a tales preguntas, por lo que la interpretación de los resultados es problemática. El desplazamiento ha sido evaluado generalmente a través de las comparaciones entre las áreas de control y las experimentales (más notoriamente en relación con la incidencia de la criminalidad). Un enfoque innovador proviene de la evaluación de las Compañías de Desarrollo Urbano (UDC) donde, usando el paralelo a la compraventa de casas en cadena,* se rastrearon los impactos más amplios de la formación de nuevos negocios mediante la identificación de lo que les sucedió a los sitios y a los locales de los negocios que se mudaron a las áreas de UDC y también mediante la clasificación de “muertes” o “nacimientos” de nuevas firmas fuera de la zona de iniciativa política (Robson et al., 1999). Quizás el ejemplo clásico del análisis del desplazamiento es el trabajo hecho sobre Zonas Empresariales (por ejemplo, Tym, 1984) —áreas en las que los incentivos fiscales fueron ofrecidos a las compañías para establecerse. Esto sugirió que un número significativo de inversiones en el interior fueron causados por las mudanzas de distancias pequeñas cruzando fronteras, lo que no representa ganancia neta alguna a la más amplia economía de la ciudad.

Un tercer problema es la necesidad de evaluar los resultados en vez de los procesos. La dificultad para medir los resultados es en parte debido a la falta de datos de áreas pequeñas que puedan ser rastreados con el tiempo. Idealmente, un estudio de los resultados necesitaría revisar indicadores en un alto nivel (de medidas tales como el desempleo, la pobreza, la migración neta o los precios de la vivienda) en una gama de escalas desde vecindarios hasta regiones de las ciudades y hacer comparaciones de los indicadores idénticos durante periodos relativamente largos. Esto proporcionaría la base para observar no sólo los cambios sustantivos en los resultados, sino también en los temas de

* N.T.: “housing chain” es un término usado para una cadena de compraventas de vivienda entre varias personas que cambian de casa, para acordar tanto la venta de la casa que dejan como la compra de la casa a la que se mudan.

desplazamiento. Como se ha señalado, la ausencia de dichos datos hasta hace poco ha hecho que dicha evaluación sea sumamente difícil. También existe, sin embargo, una segunda dificultad: que los impactos de las intervenciones de políticas están difundidos durante periodos largos, a menudo mucho después de la conclusión formal de un programa específico. Esto es tal vez lo más común en la dimensión de salud, donde los cambios significativos sólo pueden ser esperados en las generaciones siguientes (Curtis *et al.*, 2002).

El cuarto problema es el de esclarecer la causalidad en donde una variedad de intervenciones de políticas tienen lugar en una área limitada. Esto es claramente un componente esencial de la cuestión de qué es lo que funciona y de evaluar el valor por dinero. Probablemente es justo decir que esto sigue siendo el acertijo clave en la evaluación. En parte es una dificultad operacional; es decir, de desensamblar los efectos de los gastos de múltiples programas. En parte, es un problema conceptual; o sea de desenredar las líneas de la causalidad. Los programas dirigidos a un campo político podrían tener consecuencias inesperadas en otros: la generación de empleo puede ayudar a reducir las tasas de criminalidad; las mejoras de la vivienda también pueden mejorar la salud física y mental. En la medida en que las iniciativas de regeneración se convierten en multiusos y abordan una amplia gama de problemas socioeconómicos simultáneamente, los acertijos de la evaluación se han vuelto más grandes. Sin embargo, si el objetivo de la política es cada vez más—y con justa razón— el de encarar los problemas interconectados, se refuerza el argumento para usar indicadores de resultados de alto nivel para evaluar los cambios generales en los resultados y para complementar estas evaluaciones de “caja negra” con evaluaciones más suaves de proyectos específicos para rastrear las relaciones entre las actividades en el campo y los efectos en las circunstancias sociales y económicas.

Direcciones futuras

Efectivamente, esta mezcla de evaluaciones referidas a procesos con evaluaciones de resultados más cuantitativos (más duras) parece ser la meta a la que la evaluación de ABIs debiera aspirar ahora. Actualmente, los enfoques cuantitativos “tradicionales” más formales para valorar las evaluaciones de los resultados basados en metodologías cuasi-experimentales tienden a escasear, en comparación con las evaluaciones más cualitativas de procesos y estructuras. En parte, ésta es una consecuencia inevitable del énfasis a asociarse como un medio para proporcionar una regeneración más coordinada y sostenible. Esto ha ido exhortando a las evaluaciones a observar los procesos—y por ende a usar métodos cualitativos ascendentes— en lugar de los resultados medibles. Así, por ejemplo, el enfoque para observar el Presupuesto Único para la Regeneración (Rhodes, 2002) y las Compañías de Regeneración Urbana (Parkinson y Robson, 2000; AMION, 2001) ambos enfocados intensamente en la fuerza de las asociaciones. En

parte, el énfasis sobre la evaluación cualitativa también ha sido relacionado con el énfasis creciente sobre la dimensión social de la regeneración. Aun así, como Armstrong *et al.* (2002) exponen, todavía existen argumentos convincentes para aplicar enfoques cuantitativos “tradicionales” en las evaluaciones formales de resultados. Dichas evaluaciones han sido desarrolladas más convincentemente al observar los impactos económicos pequeños a nivel regional (por ejemplo, Moore y Rhodes, 1973) que al evaluar las iniciativas basadas en vecindarios más generales. Armstrong *et al.*, argumentaron que dichos enfoques son tan relevantes para las iniciativas que tienen objetivos sociales, como también para aquéllas con objetivos simplemente económicos. Los autores ilustran lo anterior con ejemplos de iniciativas de desarrollo económico comunitario (CED), las cuales tienen objetivos sociales más amplios, además de sus objetivos económicos. Ellos muestran que los esquemas del CED presentan exactamente la misma gama de acertijos para los métodos de evaluación tradicionales: objetivos múltiples; grupos de beneficiarios múltiples; la medición de la capacidad de construcción de la comunidad y los efectos que derivan de traslapar iniciativas. Los enfoques cuantitativos tradicionales todavía tienen méritos en cuanto a proporcionar una opinión cuantitativa descendente de los efectos de los programas con un énfasis social más amplio.

Dicho esto, en uno de los mensajes más contundentes que surgen de evaluar la regeneración es que son los elementos menos tangibles los que a menudo son la clave para el logro de la regeneración sustentable exitosa. El liderazgo, la calidad de individuos clave en organismos relevantes, la sensibilidad en el manejo de asuntos de la comunidad, el proceso de aprendizaje en cuanto al desarrollo de las sociedades entre las agencias; todos estos son ingredientes esenciales para el logro de una regeneración sostenible a largo plazo. Lo anterior puede ser evaluado con el uso prudente de entrevistas, discusiones, grupos focales y paneles, entre otros instrumentos. Si éstos son insumos importantes del proceso, es igualmente verdadero que un elemento esencial de los resultados lo es el factor de “sensación de bienestar”, es decir, el grado en el que las personas oriundas del lugar se sientan más seguras, más confiadas en las perspectivas de su área, más comprometidos con su futuro y con el éxito de los programas de regeneración. Mucho puede hacerse sobre los resultados duros y procesos medibles de la regeneración de nuevos empleos que puedan beneficiar sólo marginalmente a las personas locales, nuevos edificios e instalaciones que no puedan ser fácilmente accesibles debido a los costos de entrada o roces sociales, nuevos ambientes que pudiesen deteriorarse rápidamente si no hay suficiente mantenimiento ya sea por autoridades públicas o privadas.

Efectivamente, dos de las posibles direcciones que pueda tomar la evaluación futura en Gran Bretaña son precisamente abordar con mayor atención algunas de estas preguntas más sutiles. Esto es evidente tanto en la reaparición de proyectos

que están basados en investigación de acción como en la gran atención que ahora se ha otorgado a los indicadores de percepción. Un excelente ejemplo de investigación de acción es el proyecto a dos años dirigido a observar la coordinación de iniciativas diferentes en las seis localidades de East London, Plymouth, Newcastle, South Yorkshire, Sandwell y West Cumbria (Stewart et al., 2002). Esto apunta a respaldar a las diversas sociedades responsables de nueve iniciativas diferentes, muchas de las cuales se enfocan en áreas traslapadas. Gran parte de la preocupación es impulsar el trabajo conjunto para desarrollar estrategias más claras, para incentivar el intercambio de ideas y de información, y para realinear el abastecimiento de los servicios en las áreas respectivas. La evaluación más importante del Nuevo Trato a Comunidades (NDC) incorpora igualmente un elemento importante de investigación de acción.

El segundo impulso es ejemplificado por el uso a gran escala de encuestas en la evaluación actual del NDC (a pesar del costo implícito en dicho trabajo) y en el uso creciente de una serie de indicadores de Desempeño de Mayor Valor, que el gobierno está ensamblando como parte de su programa de Mayor Valor. En dicho programa se hacen preguntas de los residentes acerca de la utilización de los servicios y de su satisfacción con la calidad de los servicios, así como también acerca de la naturaleza de la entrega de los servicios —por ejemplo, los horarios de atención al público de bibliotecas, o el de accesibilidad a los edificios para los discapacitados. Pero, con el creciente interés por el papel que desempeña el “capital social” local, uno de los desafíos adicionales que los evaluadores tendrán que abordar es la mejor manera de medir este concepto. El Ministerio del Interior, por ejemplo, está desarrollando medidas de participación en la actividad voluntaria basada en la comunidad.

Si, al final del día, la regeneración y el desarrollo local son preocupaciones por hacer las áreas más atractivas para los residentes y los inversionistas, son estos conceptos más sutiles los que están en el corazón del logro sostenible a largo plazo de dicho cambio. Como Solesbury (2002) sugiere, la pregunta ha cambiado de “qué funciona” a “qué funciona para quién, bajo qué circunstancias, y mediante qué agencias”. Para enfrentar esto, es probable que lo que demuestre ser apropiado es una mezcla de metodologías de evaluación descendentes cuantitativas y ascendentes cualitativas.

Referencias bibliográficas

- AMION CONSULTING (2001), *Urban Regeneration Companies: Learning the lessons*, Department of the Environment, Transport and the Regions, Londres.
- ARMSTRONG, H. W., B. KEHRER, P. WELLS y A.M. WOOD (2002), “The Evaluation of Community Economic Development Initiatives”, *Urban Studies*, 39, 457-81.
- ATKINSON, R. y K. KINTRER (2001), “Disentangling Area Effects: Evidence from deprived and non-deprived neighbourhoods”, *Urban Studies*, 38, 2277-98.

- AUDIT COMMISSION (1989), *Urban Regeneration and Economic Development: The local government dimension*, HMSO, Londres.
- AUDIT COMMISSION (2002), *Neighbourhood Renewal*, Audit Commission, Londres.
- BEGG, I. (ed.) (2002), *Urban Competitiveness: Policies for dynamic cities*, Policy Press, Bristol.
- BENNETT, T (1996), "What's New in Evaluation Research? A note on the Pawson and Tilley article", *Bristol Journal of Criminology*, 36, 567-73.
- BRAMLEY, G., M. EVANS, J. ATKINS *et al.* (1998), *Where Does Public Spending Go? Pilot study to analyse the flow of public expenditure into local areas*, Department of the Environment, Transport and the Regions, Londres.
- CDP (1997), *Gilding the Ghetto: The state and the poverty experiments*, Community Development Project, Londres.
- CURTIS, S., B. CAVE y A. COUTTS (2002), "Is Urban Regeneration Good for Health? Perceptions and theories of the health impact of urban change", *Environment and Planning, C*, 20, 517-34.
- DORLING, D. *et al.* (2001), "How Much Does Place Matter?", *Environment and Planning, A*, 33, 1335-69.
- DETR (1998), *Urban Development Corporations: performance and good practice*, Department of the Environment, Transport and the Regions, Londres.
- DETR (2000), *Our Towns and Cities: The future. Delivering an urban renaissance*, CM4911, Department of the Environment, Transport and the Regions, Londres.
- DTLR (2001), *Changing Fortunes: Geographic patterns of income deprivation in the late 1990s*, Department of the Environment, Transport and the Regions, Londres.
- FALK, N. (2002), *Towns and Cities: Partners in urban renaissance*, URBED, Londres [www.urbed.com].
- HALL, S. y B. NEVIN (1999), "Continuity and Change: A review of English regeneration policy in the 1990s", *Regional Studies*, 33, 477-91.
- HIGGINS, J., J. DEAKIN, J. EDWARDS y M. WICKS (1983), *Government and Urban Policy: Inside the policy-making process*, Blackwell, Oxford.
- HM TREASURY (1995), *A Framework for the Evaluation of Regeneration Projects and Programmes*, HMSO, Londres.
- HO, S. Y. (1999), "Monitoring and Evaluation of British Urban Policy: Case study of City Challenge", *Tesis doctoral*, University of Manchester.
- KPMG CONSULTING (1999), *City Challenge: Final national evaluation*, Department of the Environment, Londres.
- LAWLESS, P., G. DABINETT, P. TYLER y J. RHODES (2000), *A Review of the Evidence Base for Regeneration, Policy and Practice*, Department of the Environment, Transport and the Regions, Londres.
- LENIHAN, H. (1999), "An Evaluation of Regional Development Agency's Grants in Terms of Deadweight and Displacement", *Environment and Planning, C*, 17, 3003-18.

- MACLENNAN, D. (2000), *Changing Places, Engaging People*, York Publishing Services, York.
[Para detalles sobre las publicaciones de Joseph Rowntree del Programa para la Regeneración de Áreas, véase www.jrf.org.uk].
- MCCULLOCH, A. (2001), "Ward-level Deprivation and Individual Social and Economic Outcomes in the British Household Panel", *Environment and Planning, A*, 33, 667-84.
- MOORE, B. y J. RHODES (1973), "Evaluating the Effects of British Regional Policy", *Economic Journal*, 83, 87-110.
- NOBLE, M. et al. (2000), *Measuring Multiple Deprivation at the Small Area Level: The Indices of Deprivation 2000*, Department of the Environment, Londres.
- PARKINSON, M. P. y B.T. ROBSON (2000), *Urban Regeneration Companies: A process evaluation*, Department of the Environment, Transport and the Regions, Londres.
- PAWSON, R. y N. TILLEY (1994), "What Works in Evaluation Research?", *British Journal of Criminology*, 34, 291-306.
- REGIONAL CO-ORDINATION UNIT (2002), *Review of Area Based Initiatives*, Regional Coordination Unit, Office of the Deputy Primer Minister, Londres.
- RHODES, J., P. TYLER, A. BRENNAN, S. STEVENS, C. WARNOCK, M. OTERO-GARCÍA (2002), *Lessons and Evaluation Evidence from Ten Single Regeneration Budget Case Studies*, Department of the Environment, Transport and the Regions, Londres.
- ROBSON, B.T., M.G. BRADFORD y I. DEAS (1999), "Beyond the Boundaries: Vacancy chains and the Urban Development Corporations", *Environment and Planning, A*, 31, 647-64.
- ROBSON, B.T., M.G. BRADFORD, M. PARKINSON et al. (1994), *Assessing the Impact of Urban Policy*, Department of the Environment, HMSO, Londres.
- ROBSON, B.T., M.G. BRADFORD y R. TOMLINSON (1998), *Updating the Index of Local Deprivation*, Department of the Environment, Transport and the Regions, Londres.
- ROBSON, B.T., M.G. BRADFORD y R. TYE (1995), "A Matrix of Deprivation in English Authorities, 1991" In *Deprivation Index: A review of approaches and a matrix of results*, Department of the Environment, HMSO, Londres, pp. 69-163.
- ROSSI, P. H., H.E. FREEMAN y M.W. LIPSEY (1999), *Evaluating: A systematic approach*, 6th. ed. Sage Publications, Thousand Oaks, California.
- RUSSELL, H., J. DAWSON, P. GASROIDE y M. PARKINSON (1996), *City Challenge: Interim national evaluation*, Department of the Environment, Londres.
- SOCIAL EXCLUSION UNIT (2001), *A New Commitment to Neighbourhood Renewal: National Strategy Action Plan*, Cabinet Office, Londres.
- SOLESBURY, W. (2002), "Evidence Based Policy: Whence it came and where it's going", *Planning Policy and Practice*.

STEWART, M. et al. (2002), *Collaboration and Co-ordination in Area-Based Regeneration Initiatives*, Department of the Environment, Transport, Local Government and the Regions, Londres.

TYLER. P., J. RHODES y A. BRENNAN (1998), *The Distribution of SRB Challenge Fund Resources in Relation to Local Area Needs in England*, Department of the Environment, Transport and the Regions, Londres. [Véase también A. Brennan, J. Rhodes y Tyler (1999) *Urban Studies*, 36, 2069-84].

TYM, R. y PARTNERS (1984), *Monitoring Enterprise Zones: Three year report*, Department of the Environment, Londres.

Capítulo 9

Comentario sobre el artículo y taller “Evaluación de las políticas de áreas” de Brian Robson

Por
Jonathan Potter,
Programa OCDE LEED

Este comentario se concentra en tres puntos principales del artículo del profesor Robson y la discusión del taller sobre la experiencia del Reino Unido en la evaluación de la política de áreas. Dichos comentarios contienen asuntos que también son importantes para la práctica de evaluación en otros países. En primer lugar, la calidad y utilidad de la evaluación de la política de áreas son a menudo menores de lo que podrían ser debido a que ciertos conceptos básicos de evaluación no son abordados sistemáticamente. En segundo lugar, debemos reconocer que existen varias barreras para realizar evaluaciones con calidad de la política de áreas en el ámbito local, lo que limita el grado de aplicación de los conceptos clave para evaluar. En tercer lugar, existen diversas formas en que los gobiernos municipales y las agencias de desarrollo pueden apoyar para incrementar la capacidad de evaluación en el ámbito local, ayudando —de esta forma— a vencer los obstáculos para una más rigurosa evaluación de la política de áreas.

Conceptos clave para evaluar con calidad la política de áreas

Tal y como el trabajo de Profesor Robson apunta, las tres décadas anteriores han registrado un incremento en la evaluación políticas de área en el Reino Unido. Adicionalmente, la llegada del gobierno laborista en 1997 trajo una expansión importante en el trabajo de evaluación de este tipo de políticas, debido al incremento de la gama de programas de área y a un compromiso de formulación de políticas con mayor base en evidencias a través del gobierno en su conjunto. Sin embargo, la evaluación que se ha llevado a cabo refleja una calidad variable. Mientras Robson señala algunas de las mejores prácticas de las evaluaciones, también argumenta que muchas evaluaciones a pequeña escala han fallado al lidiar en la práctica con ciertos asuntos conceptuales que deben ser incluidos en cualquier evaluación rigurosa de la política de áreas. Algunas cuestiones que requieren más atención en la práctica actual de la evaluación son esbozadas a continuación en una idea general y en partes, ampliando los argumentos del trabajo de Robson:

- *Razón fundamental y objetivos del programa.* Una evaluación debe empezar por comprender la razón de ser y los objetivos originalmente previstos para un

programa, i. e. qué problemas esperaba abordar el programa y cómo se esperaba que los encarara. La razón fundamental subyacente puede estar incorporada en la documentación de la política, pero a menudo dicha razón debe ser extraída por el evaluador mediante entrevistas con los diseñadores de la política local. El evaluador debe valorar si la razón fundamental y los objetivos fueron apropiados y si éstos todavía son apropiados, y también comparar los resultados del programa con la razón fundamental y con los objetivos para valorar la eficacia de programa en términos de los mismos. Un problema que surge a menudo es que la razón fundamental detrás de la política resulta estar perdida en mayor o menor grado. Por lo tanto, se han establecido múltiples discusiones de la tendencia de los diseñadores de política a aplicar "recetas generales" o seguir las modas pasajeras, sin considerar completamente si el enfoque es realmente apropiado para las necesidades y oportunidades del área. Un ejemplo clásico es el de las localidades que buscan atraer la industria de alta tecnología debido a sus sueldos altos y crecimiento potencial, incluso si las condiciones locales no existen para el desarrollo sostenible del sector (enlaces con las universidades, mano de obra calificada, buenas comunicaciones, atractivo residencial, etcétera). Un asunto relacionado es la consideración de las alternativas, i.e. ¿podrían encontrarse maneras alternativas de lograr los mismos objetivos de forma más eficaz? Una evaluación rigurosa debe valorar la razón fundamental del programa considerando si una alternativa es apropiada y si es la mejor manera de cumplir los objetivos.

- *Pérdida de eficiencia social (Deadweight)*. Demasiadas evaluaciones fallan al considerar las situaciones contrafactuales; por ejemplo: ¿qué hubiera ocurrido con la ausencia de un programa? En particular, es importante considerar qué tan lejos habrían tenido lugar de todos modos las acciones alentadas por la política (pérdida de eficiencia, peso muerto). Sin tal valoración, es fácil sobreestimar los efectos del programa atribuyendo todos los resultados positivos en el área en la que el programa está siendo evaluada. Cuando el contrafactual ha tomado lugar en las evaluaciones del Reino Unido, los evaluadores han tendido a depender de valoraciones personales hechas por los diseñadores de la política encuestados, directores del programa, beneficiarios y otras partes interesadas. Es difícil tener confianza plena en tales pruebas porque los encuestados encuentran extremadamente complejo responder a la hipotética pregunta "¿qué hubiera sido?" y porque el interés personal puede conducir a respuestas sesgadas. Grupos de control o comparaciones con áreas de control pueden complementar los ejercicios de valoraciones personales y proporcionar al evaluador mayor confianza en los resultados.
- *Desplazamiento y sustitución*. El desplazamiento ocurre donde el inicio de operaciones o el crecimiento de empresas y organizaciones alentadas por la política resulta en una pérdida de actividad en otras empresas y organizaciones.

En el contexto de políticas en el mercado de trabajo, la sustitución ocurre cuando la política conduce a que las personas apoyadas obtengan empleos a expensas de los individuos no apoyados. A menudo estos efectos no son considerados completamente por los evaluadores. Como con el peso muerto, los evaluadores dependen de los ejercicios de valoración personal sobre el grado en que otros individuos, empresas o áreas han sido probablemente afectados negativamente. Las comparaciones con grupos y áreas de control deberían ser usadas más frecuentemente para complementar esta información. También ha sido relativamente poco considerado hasta la fecha, si el desplazamiento y la sustitución deben ser descontados completamente como efectos negativos o si pueden ser considerados en parte como un estímulo positivo del ajuste dinámico y de la innovación de la economía local. Esto último podría ser verdadero si la política da como resultado el desarrollo de ventajas competitivas en las organizaciones y las personas sobre otros, que dan como resultando una mayor productividad en general. Además, ha sido argumentado que la simple creación del “nerviosismo” en el mercado laboral puede ser beneficioso si esto protege a las personas del daño de estar inactivas durante mucho tiempo. Por lo tanto, el desafío para los evaluadores es medir en primer lugar el grado de desplazamiento y sustitución y, en segundo lugar, determinar qué tanto esto debe ser considerado como un fenómeno positivo o negativo.

- *Evaluaciones longitudinales.* Las evaluaciones de política de área han sido a menudo limitadas a valoraciones *ex post* al terminar el programa. Pero es difícil explorar qué tan bien ha funcionado el programa si la evaluación solamente tiene lugar después del evento debido a los cambios que tienen lugar durante el programa, como las inauguraciones y los cierres de firmas, el movimiento de las personas y los cambios en la naturaleza de las actividades apoyadas. Algunas evaluaciones en el Reino Unido han intentado desarrollar un enfoque longitudinal, generalmente a través de la elaboración de análisis de corte transversal de la actividad dentro del área en dos o más puntos en el tiempo. Por ejemplo, las encuestas anuales sobre el uso del suelo y de actividad de las empresas en las zonas de Corporaciones de Desarrollo Urbano del Reino Unido ayudaron a indicar los procesos de cambio durante la vida de estas iniciativas. Es importante seguir de este modo la trayectoria de los programas en el tiempo.
- *Movilidad de los hogares.* Cuando al revisar qué tanto la política está ayudando a las personas que viven en barrios marginados, las evaluaciones deben enfrentar también el asunto complejo de los efectos de la movilidad de los hogares en el interior y hacia afuera del área objetivo en los resultados de la evaluación. La dificultad aparece porque las personas apoyadas pueden mudarse fuera de la zona y los beneficios asignados a ellos no serían atribuidos a la iniciativa a menos que la evaluación tuviera alguna manera de rastrear a las personas

apoyadas. Por otra parte, una iniciativa de regeneración puede alentar a las personas a mudarse a la zona, mejorando el perfil de población del área objetivo y dar la impresión que la política ha ayudado a las personas de la localidad si las personas entrantes son relativamente más adineradas, aunque la política no podría haber ayudado directamente a la población objetivo original. En el Reino Unido, la evaluación de la política de áreas rara vez ha tratado de seguir el movimiento de las personas que la política intenta ayudar, pero esto debe ser considerado en el diseño de evaluaciones rigurosas de políticas de áreas.

- *Los resultados (beneficios e impactos) versus los productos (acciones y cifras).* Muchas evaluaciones se detienen en cuantificar las acciones y cifras de un programa, por ejemplo, el número de personas que obtienen alguna aptitud, el número de personas que encuentran empleo, o el número de hectáreas de regiones abandonadas otra vez en uso. Sin embargo, lo que realmente interesa son los impactos y beneficios resultantes en términos de la mejora de la calidad de vida de las personas y de la estabilización o mejora del dinamismo económico y social de la zona. Por consiguiente, el valorar los resultados (beneficios e impactos) debe ser la meta final de la evaluación de la política de áreas, concentrándose en medidas tales como ingresos, pobreza, salud, migración y precios del metro cuadrado de los terrenos.
- *Esclareciendo la causalidad.* Establecer la causalidad es especialmente importante para las políticas de áreas, porque éstas tienden a poner en un lugar múltiples intervenciones que abordan varios aspectos económicos, sociales y de regeneración física. La complicación para la evaluación es que las acciones dirigidas a una actividad pueden tener impactos sobre otras. Por ejemplo, las acciones para la generación de empleo pueden reducir el crimen, mientras que las medidas para reducir el crimen pueden impulsar la creación de actividades mercantiles. Para juzgar la combinación correcta de políticas para una zona, los diseñadores de políticas necesitan saber la contribución relativa de las diferentes acciones al total de los productos colocados y de los resultados obtenidos (impactos, beneficios). Todavía hay mucho que progresar para medir el grado y la naturaleza de las sinergias entre las acciones. Algunas veces los enfoques matriciales son aplicados y muestran si determinadas acciones contribuyen o si es probable que contribuyan en las mejoras en más de un ámbito. Esto provee una estructura para identificar las sinergias posibles, pero no va muy lejos en la medición del grado cómo éstas ocurren. La necesidad de medir las sinergias es cada vez más acreditada, pero es necesaria mayor investigación para desarrollar técnicas más sofisticadas para hacerlo.
- *Datos de áreas pequeñas.* Debido a que la política de área se dirige a zonas geográficas pequeñas y los datos oficiales tienden a estar disponibles sólo para unidades administrativas mayores, los evaluadores enfrentan con frecuencia problemas para obtener datos apropiados para medir la situación del punto de

partida y los cambios subsecuentes en el área comparando con los registrados en áreas vecinas y en la economía más amplia. Como Robson apunta, este asunto ha sido reconocido en el Reino Unido y un esfuerzo considerable se está realizando para mejorar dicha situación.

- *Escala de intervención.* Las evaluaciones de políticas de área tienden a hacer su referencia central a la comparación entre los resultados o impactos del programa con los gastos del programa. Hay dos problemas principales con esta manera de pensar. En primer lugar, existe la tendencia de no ponderar si la política está siendo aplicada en la escala suficiente para abordar los problemas en un lapso de tiempo satisfactorio. Una pregunta clave que deben hacerse los evaluadores es, "si esta política es aplicada a esta tasa, ¿cuánto tiempo tomaría resolver los problemas del área?" En segundo lugar, existe también la tendencia de no ponderar si, considerando un incremento en los recursos, una iniciativa eficaz y eficiente de pequeña escala podría ser reescalada para enfrentarse con áreas y grupos más amplios. Aparte del asunto de si cualquier programa funcionaría del mismo modo en un contexto diferente, la omisión de considerar la escala puede resultar en recetas de política engañosas que naturalmente sugieren que los programas exitosos deben ser ampliados. Los esquemas a pequeña escala frecuentemente parecen muy prometedores porque pueden tratar con las áreas y los grupos en los cuales la política probablemente hace mayor diferencia, pero podría ser difícil lograr el mismo éxito con áreas y grupos más grandes.

¿Por qué la evaluación con calidad de la política de áreas es relativamente escasa?

Los participantes del taller trataron de establecer por qué la evaluación con calidad de la política de áreas que toma en cuenta los aspectos conceptuales referidos arriba es relativamente infrecuente. Las siguientes barreras posibles fueron identificadas:

1. La evaluación involucra muchas externalidades positivas que son difíciles de adjudicarse para un gobierno municipal o una agencia de desarrollo. Por ejemplo, frecuentemente los resultados son hechos públicos, por lo que es fácil para las otras áreas "aprovecharse" de las evaluaciones realizadas por otros mediante referencias.
2. Existen economías de escala en la evaluación, las cuales indican que dicha actividad puede ser mejor emprendida en el ámbito central o por grupos de áreas locales que cooperen conjuntamente, en vez de que las realicen los gobiernos municipales u organismos en lo individual.
3. Una evaluación de buena calidad puede ser sumamente costosa y por lo tanto puede llegar a significar una proporción importante del gasto total del

programa, especialmente en programas pequeños. En términos económicos, esto puede pensarse como un problema de no divisibilidad en lo que respecta a que los gastos de la evaluación rigurosa son difíciles de reducir incluso para programas pequeños. Por lo tanto organismos locales tienen que hacer un balance entre los méritos relativos de ampliar un programa o evaluar completamente el programa original. Probablemente no es inteligente gastar grandes cantidades de dinero en evaluar un programa muy pequeño.

4. En los Estados Unidos, aunque menos que en Europa y en muchos otros países de la OCDE, la mayoría de los programas de desarrollo económico son financiados a través de renuncias a las exenciones fiscales en lugar del gasto público. Así, el gasto de las iniciativas de desarrollo económico es menos visible y a menudo no es percibido como un costo real por los funcionarios y el electorado. Parece ser menor la presión para evaluar programas financiados a través de renuncias a las exenciones fiscales aunque los grupos de presión en los EU están empezando a poner atención en este asunto.
5. Siguiendo una lógica similar, generalmente el desarrollo económico llevado a cabo a nivel local no es financiado localmente sino nacionalmente, regionalmente o a través de otros fondos. Donde se gastan pocos recursos locales, es poco probable que los gobiernos municipales y organismos locales deseen evaluar un programa aunque su repercusión sea local.
6. Existe una falta de sincronización fundamental de los marcos temporales entre el desarrollo político y económico. Así, el marco temporal político tiende a ser un ciclo electoral, mientras que el desarrollo económico puede requerir de una o dos décadas. Es probable que los políticos locales estén reacios para autorizar evaluaciones importantes a largo plazo si ello aportara poco en la forma de evidencia que apoye sus actividades en curso.
7. Muchos profesionales del desarrollo local parecen estar satisfechos con hacer referencia a ejemplos de estudios de caso de las mejores prácticas en lugar de ordenar la elaboración y usar sus propias evaluaciones. Esto podría reflejarse el costo más bajo del intercambio de las mejores prácticas y en el atractivo de mantenerse al día con las nuevas tendencias presentadas en conferencias y en publicaciones. Pero, mientras que el intercambio de las mejores prácticas pueda ser útil, especialmente cuándo está basado en información de programas evaluados en su totalidad, dicho intercambio no es un sustituto de las evaluaciones directas a programas locales. Aquí el problema parece ser una falta de consumidores informados para impulsar la evaluación con calidad de la política de áreas en el ámbito local.

Construyendo capacidades en la evaluación de políticas de áreas

En respuesta a las hipótesis anteriores, los participantes del taller sugirieron ciertas acciones que podrían incrementar la cantidad y la calidad de la evaluación de políticas de área y su uso por los funcionarios de políticas locales y por políticos en el diseño e implementación de los programas futuros.

- Alentar los mecanismos que conduzcan al suministro local de recursos financieros de programas, tales como las concesiones en bloque del gobierno central a los gobiernos municipales y organismos locales con lo que pueden tener flexibilidad en la elección y en el diseño de programas para sus zonas o programas y les permiten utilizar los impuestos captados y retenidos en la localidad para los propósitos de regeneración. El financiamiento local de programas generará presión local para evaluar ese gasto y resultará en un mayor sentido de propiedad de los hallazgos y de compromiso para utilizarlos.
- Impulsar las evaluaciones emprendidas en colaboración con los grupos de gobiernos municipales o agencias de desarrollo o apoyados por agencias de más alto nivel con el objeto de abordar los problemas de las externalidades, las economías de escala y el mayor costo de evaluación representado como el porcentaje del presupuesto del programa.
- Hacer frente a las diferencias entre las escalas de tiempo de los evaluadores, de los políticos y de los diseñadores de políticas para asegurar mayor demanda por la evaluación y por el uso de ésta. Esto significa hacer hincapié en los políticos y en los diseñadores de la política quienes deben comprometerse con el desarrollo económico como un ejercicio de largo plazo, por ejemplo, disponiendo la perspectiva de las agencias de desarrollo de acuerdo con las metas de largo plazo, y estableciendo que no pueden esperar a que los evaluadores demuestren el éxito inmediato. Al mismo tiempo, debe hacerse hincapié en que los evaluadores políticos y los diseñadores de políticas necesitan resultados intermedios para ayudar a justificar la continuación de programas que parecen estar funcionando o para ajustar los programas que aparentan no estar desempeñándose bien.
- Similarmente, hacer frente a las diferencias de lenguaje entre evaluadores, políticos y diseñadores de políticas, para incrementar otra vez la demanda y uso de la evaluación, aumentar el conocimiento de los políticos y de sus asesores acerca de las técnicas de evaluación y sus implicaciones y simplificar la manera en que los evaluadores presentan sus hallazgos.
- Mejorar las capacidades de los diseñadores de políticas locales para comisionar y utilizar las evaluaciones de la política de área, incluyendo el desarrollo de la habilidad sobre cómo juzgar la calidad del trabajo de evaluación y cómo trasladarlo a la próxima fase de mejoramiento de la política y la práctica.

Tal capacitación del cliente podría brindarse a través de la educación y el intercambio de información para los profesionales del desarrollo local. Los siguientes pasos específicos pueden ser sugeridos:

- Desarrollar foros en los cuales podrían intercambiarse la experiencia entre las personas que realizarán políticas de área en el futuro y utilizarán evaluaciones locales.
- Establecer centros de la excelencia para la capacitación en los métodos de política de área donde las técnicas y los usos de la evaluación pudieran ser un componente importante.
- Apoyar la difusión pública de ejemplos de evaluaciones de buena calidad, por ejemplo usando sitios en Internet, revistas y publicaciones dirigidas a los profesionales de regeneración local.

Capítulo 10

Evaluación de los programas de apoyo a negocios

por
Eric Oldsman,
Nexus Associates Inc.
y
Kris Hallerg*
Departamento de evaluación de operaciones,
Banco Mundial, Washington

* Las opiniones expresadas aquí son de los autores y no representan las políticas del Grupo del Banco Mundial.

Prólogo

Los gobiernos alrededor del mundo están apoyando una amplia gama de programas de asistencia para negocios dirigidos a promover el desarrollo de empresas privadas, en particular las empresas pequeñas y medianas (PyMEs). A pesar de la cantidad de recursos comprometidos en estos programas, ha habido poco esfuerzo dedicado a determinar si en realidad estos programas han tenido éxito en alcanzar los resultados esperados. En conjunto, las evaluaciones se han apoyado en estudios de antes-y-después con errores heredados o bien en testimonios potencialmente sesgados de usuarios agradecidos.

Sin embargo, para los gobiernos hay mejores alternativas disponibles. Las encuestas de beneficiarios potenciales claramente tienen un lugar en las evaluaciones, posibilitando a los evaluadores detectar información útil sobre la percepción de los participantes. Pero se requiere tener cuidado y asegurarse de que las encuestas deben enfocarse en los aspectos críticos del diseño del programa, que estén escritas en palabras que tomen en cuenta directamente la situación contrafactual, y que se basen en muestras representativas.

Dicho lo anterior, bajo ciertas circunstancias, el juicio de los participantes puede no proporcionar evidencia suficiente del impacto de los programas. Aquí, los gobiernos pueden preferir aplicar diseños experimentales o cuasi-experimentales más rigurosos para obtener estimados válidos del impacto de programas en particular, controlando factores ajenos que puedan influir en los cambios observados del desempeño. Técnicas estadísticas pueden ser usadas para probar varias hipótesis sobre el impacto de las variables clave. Además, dentro de lo posible, las evaluaciones deben incluir estudios de caso apoyados en narrativas detalladas para explicar mecanismos causales e identificar elementos del diseño del programa que requieran ser modificados. Finalmente, sin importar el enfoque particular, todas las evaluaciones deben estar fundadas en declaraciones precisas que detallen la población objetivo, los resultados esperados, y las suposiciones que vinculan las actividades del programa y las metas declaradas.

Este estudio busca proporcionar a funcionarios gubernamentales una mejor comprensión de los aspectos críticos involucrados en la evaluación de programas y las varias herramientas que pueden usarse en evaluaciones acumulativas de programas de apoyo de negocios dirigidos a PyMEs.

Introducción

Los gobiernos alrededor del mundo han invertido montos considerables de dinero en una variedad de iniciativas para promover el desarrollo de negocios privados, particularmente pequeñas y medianas empresas (PyMEs).¹ El interés de los gobiernos federales, estatales y locales en las PyMEs surge, en parte, de reconocer que las mismas juegan un papel crítico en todas las economías; ellas producen una gran gama de bienes y servicios para consumo doméstico y exterior y, al hacerlo así, son una fuente importante de ingresos y trabajos en cada región.

Mientras que las PyMEs constituyen una parte significativa de la economía, muchos creen que el desempeño de ellas está por debajo del óptimo desde la perspectiva social y sostienen que se requiere la intervención del gobierno para impulsar el crecimiento y rendimiento de tales empresas. Quienes son partidarios de la intervención gubernamental señalan varias imperfecciones en los mercados relevantes como justificación de dicha acción.² En algunos casos, se argumenta que los servicios requeridos por las PyMEs no están disponibles en el mercado. En otros casos, la razón fundamental ronda en que las PyMEs carecen de la información requerida para tomar decisiones correctas de compra o inversión. Incluso en otros casos, las razones se centran en la aseveración que las decisiones de empresas privadas individuales para perseguir un curso de acción en particular no reflejan beneficios o costos sociales más amplios.

Las decisiones para proveer de fondos iniciativas dirigidas a las PyMEs están basadas en la creencia que programas bien diseñados se harán cargo de las imperfecciones de mercado, impulsarán el desempeño de las compañías participantes y rendirán beneficios económicos y sociales significativos. Los gobiernos ahora buscan evidencia creíble de que estas creencias eran ciertas y que programas particulares de apoyo a negocios justifican un respaldo continuo.

La búsqueda de metodologías adecuadas de evaluaciones de los programas refleja el hecho que los gobiernos están bajo presión constante para asignar recursos escasos a necesidades concurrentes. Estas elecciones son más difíciles, aunque necesarias, en el contexto actual, en el que es muy probable que se reduzca significativamente el costo discrecional. Información sobre los resultados reales de los programas establecidos por el gobierno para enfrentar necesidades variadas es crítica para las deliberaciones presupuestarias; las evaluaciones pueden proveer bases para trasladar recursos de programas con bajo desempeño a aquellos que han demostrado ser exitosos.

Al mismo tiempo, los gobiernos necesitan la información para identificar áreas en las que se requieren cambios en el programa para mejorar las oportunidades de éxito. Los programas pueden requerir un afinamiento o ser objeto de modificaciones sustanciales basándose en evidencia sólida de lo que

funciona o no y el porqué. Para esto, la evaluación puede ayudar a identificar factores críticos de éxito y proveer una base para la toma de decisiones informadas sobre cómo rediseñar ciertos programas en particular.

Las decisiones relacionadas con la continuidad de los fondos o sobre las operaciones en marcha deben apoyarse en información exacta y creíble. Respecto a esto, las evaluaciones deben estar bien diseñadas y deben implementarse de acuerdo con buenos estándares de práctica. Como se aborda en la próxima sección, esto debe empezar con una articulación clara del diseño del programa.

Diseño del programa

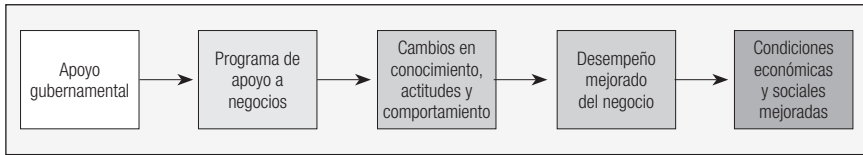
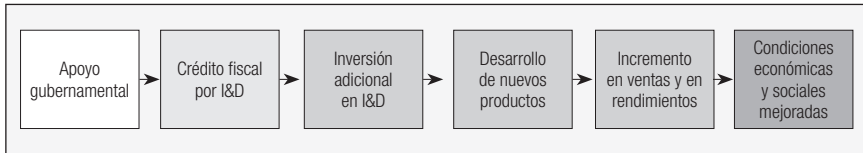
Las iniciativas de los PyMEs se dirigen a empresas (de un solo propietario, sociedades y corporaciones) dentro de rangos específicos de tamaño, como lo son: ventas anuales, empleos o activos. Algunas veces se usan otras características o condiciones para definir las PyMEs objetivo. Por ejemplo, algunos programas en particular pueden ser dirigidos a empresas propiedad de mujeres, empresas rurales o sectores específicos de la industria.

Después de alguna ponderación de las necesidades de la población objetivo, los gobiernos han instituido programas que incorporan uno o más de los elementos siguientes:

- Servicios de consultoría administrativa o técnica para enfocarse en varios procesos de negocios, incluidos: planeación, desarrollo de productos, mercadeo y ventas, producción, distribución, recursos humanos, sistemas de información y administración financiera.
- Servicios de capacitación para mejorar las habilidades de la administración, de los supervisores, los operarios de equipos u otro personal de la compañía a través de una combinación de instrucción en salón de clases y práctica.
- Concesiones y otras formas de financiamiento para inversión de capital, requerimientos de capital de trabajo u otras necesidades.
- Créditos fiscales para inversión en investigación y desarrollo, capital para equipamiento y entrenamiento de los empleados.
- Acceso a servicios de bajo costo, equipo y otra infraestructura física.

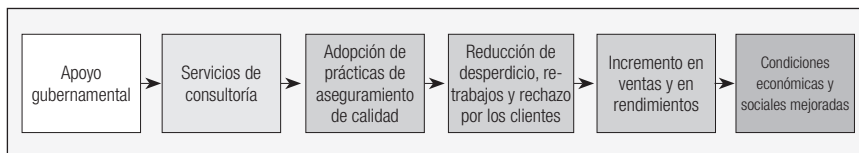
Todos los programas para PyMEs, independientemente de los elementos incluidos, están dirigidos a obtener los resultados deseados a nivel de la empresa y la economía en general. Específicamente, estos programas apuntan a cambiar el comportamiento de las empresas participantes, que resultan en el mejoramiento del desempeño del negocio y, a su vez, en el mejoramiento de las condiciones económicas y sociales, como se muestra en la figura 10.1.

Modelos más detallados de lógica para el programa pueden desarrollarse para programas específicos. Por ejemplo, como se muestra en la figura 10.2, un

Figura 10.1 **Modelo lógico de un programa básico**Figura 10.2 **Modelo lógico para PyMEs para créditos fiscales por I&D**

crédito fiscal en investigación y desarrollo (I&D) dirigido a todas las PyMEs en una jurisdicción fiscal específica tiene el propósito de proporcionar beneficios fiscales a las compañías, bajando así el costo de I&D e induciendo de este modo inversión adicional en los esfuerzos de desarrollo de productos. Esto, a su vez, se espera que conduzca al verdadero desarrollo de productos novedosos que satisfagan las necesidades del cliente y de esta forma a su subsiguiente introducción a los mercados comerciales. Además, se espera que estos productos probarán ser superiores a otros productos en el mercado, generando ventas crecientes y redituables para las compañías, así como beneficios para los consumidores. También se asume que las compañías que se benefician del crédito fiscal en I&D contratarán investigadores adicionales para hacerse cargo del programa expandido de I&D y otros empleados necesarios para cubrir la creciente demanda de los productos comerciales resultantes. A este respecto, se asume que los productos serán producidos por las compañías que se benefician directamente del crédito fiscal por I&D o licencias localizadas en la misma región o país.

En comparación, los programas de consultoría de negocios están dirigidos a proveer información a las compañías que las conduzca a efectuar cambios en sus operaciones que de otro modo no hubieran realizado, obteniendo con esto mejoras en procesos en particular y en el desempeño general en toda la empresa. Por ejemplo, como se ilustra en la figura 10.3, un programa puede enfocarse a proveer información sobre la importancia de instituir buenos procedimientos que aseguren la calidad. Se espera que esto lleve a las compañías a adoptar ciertas prácticas, como el control estadístico de la calidad o buscar obtener la certificación ISO 9000. Se anticipa que la implantación de estas prácticas resultará en una mejora de la calidad, como se evidencia por la reducción en la tasa de desperdicio, repetición de trabajos o rechazos por usuarios; a su vez, se espera

Figura 10.3 **Modelo lógico para programas de consultoría a PyMEs**

que la mejor calidad lleve a un incremento en las ventas y los rendimientos. Dependiendo de la relación entre el aumento anticipado de las ganancias en la productividad y el crecimiento en ventas, los programas pueden resultar en incrementos en el empleo.

Estos dos ejemplos demuestran como los modelos de lógica de los programas pueden proveer una descripción concisa de cómo los programas mejorarán las condiciones dentro de la población objetivo, observando mecanismos causa-efecto importantes (si X, entonces Y). Como resultado, ambos ejemplos muestran hipótesis que ostensiblemente podrían ser probadas en las evaluaciones de los programas. Por ejemplo, en el caso del crédito fiscal para I&D la hipótesis clave es que el crédito conduce a inversiones en I&D que de otra forma no hubieran sido implementados por la empresa.³ De manera similar, para el programa de consultorías, la hipótesis principal es que el servicio conduce a acciones específicas que de otra forma no habrían sido llevados a cabo por las empresas.

Por lo tanto, antes de embarcarse en una evaluación, el o los objetivos del programa, así como el camino que conecta las actividades al resultado esperado, deben ser definidos tan claro como sea posible.⁴ El modelo lógico del programa resultante debe ser usado para definir el ámbito de la evaluación, identificar resultados que deben ser medidos y ayudar a proveer la base para afirmar la causalidad.

Indicadores de resultados

Con un modelo lógico para el programa, el siguiente paso es establecer un conjunto de indicadores mensurables que puedan ser usados para constatar el impacto de un programa de asistencia a un negocio en particular. Al desarrollar estas medidas, es esencial tomar en consideración lo siguiente:

- **Relevancia.** Las medidas seleccionadas para la apreciación del impacto deben guardar relación con la iniciativa en particular a ser estudiada.
- **Validez.** Se requiere que las medidas provean un reflejo preciso del concepto subyacente que se supone es medido.

- **Confiabilidad.** Las medidas deben estar sujetas al menor error de medida posible.
- **Conveniencia.** Debe ser posible obtener los datos necesarios para efectuar la medición.

El resultado de las evaluaciones sólo será preciso y creíble en el punto en que las medidas sean relevantes, válidas y confiables. Pero también debe ser factible utilizar medidas basadas en la disponibilidad de datos, de tiempo, y de limitaciones presupuestarias. Por ejemplo, existe una variedad de maneras para cuantificar la productividad; *v. gr.* producción por empleado, valor agregado por hora de trabajo, productividad total de los factores. La última medida refleja el valor adicional generado mediante el uso del capital, el trabajo, los materiales y otros factores de producción. Aunque se puede decir que es la mejor medida de la productividad, es muy difícil obtener los datos requeridos, incluso en compañías grandes con sofisticados sistemas de información. Por otro lado, aunque la producción por empleado como medida de productividad puede ser engañosa dado que el incremento en la contratación externa (*outsourcing*) se muestra como una ganancia de productividad, es relativamente fácil obtener los datos necesarios. Al hacer un balance, ésta puede ser la mejor elección para una apreciación de un impacto específico. Como otros aspectos del diseño e implementación de la evaluación, la selección de medidas de resultados frecuentemente involucra sacrificios.

Las medidas de resultados requieren ser desarrolladas dentro del contexto de iniciativas en particular, reflejando las metas y objetivos específicos de la intervención, así como las preocupaciones prácticas respecto a la disponibilidad de datos. No existe un conjunto de medidas que puedan adecuarse a toda iniciativa de programa de asistencia a negocios, pero puede haber indicadores similares para programas similares. Se dan algunos ejemplos en la tabla 10.1. Muchos de estos indicadores se enfocan a los cambios de calidad, tiempo de rotación y costos de producción. Otros tienen el propósito de medir el desempeño global de la empresa respecto a cambios en ventas, ganancias netas y empleo.

Tabla 10.1 **Indicadores potenciales de resultados para PyMEs señaladas**

<i>Indicador</i>	<i>Definición</i>
Cambios de actitud	Persistencia e incidencia de actitudes entre administradores, supervisores o trabajadores.
Cambios de procesos	Persistencia e incidencia de cambios en procesos específicos, <i>v. gr.</i> planeación, ventas y mercadeo, producción y distribución.
Inversión	Montos invertidos en planta, equipo, <i>software</i> o entrenamiento.
Nivel de defectos (repetición de trabajos o desperdicio)	Proporción de las unidades que no se adecuan al diseño estándar y subsecuentemente son trabajadas de nuevo o desechadas.

Tabla 10.1 **Medidas potenciales de resultados para PyMEs señaladas** (cont.)

<i>Indicador</i>	<i>Definición</i>
Tiempo entre la orden y la entrega	Tiempo total (horas o días) desde la recepción de la orden hasta su entrega bajo las premisas del cliente.
Entregas a tiempo	Proporción de las órdenes entregadas a los clientes en el tiempo programado.
Rechazos de clientes	Proporción de ítems entregados a los clientes y subsecuentemente rechazados por inconformidad.
Utilización de la capacidad	Proporción de los recursos disponibles (<i>v. gr.</i> planta y equipo) usada en la producción.
Productividad de la mano de obra	Valor de venta de la producción del periodo dividido entre las horas de mano de obra directas usadas en la producción.
Ventas	Ganancias derivadas de la venta de bienes y servicios.
Ganancia neta	Ganancia de operación (ventas menos costo de los bienes vendidos) y otros ingresos menos gastos totales.
Empleos	Trabajadores de tiempo completo o parcial empleados por las compañías o propietarios únicos a una fecha específica o periodo de pago, <i>v. gr.</i> la semana del 12 de marzo.

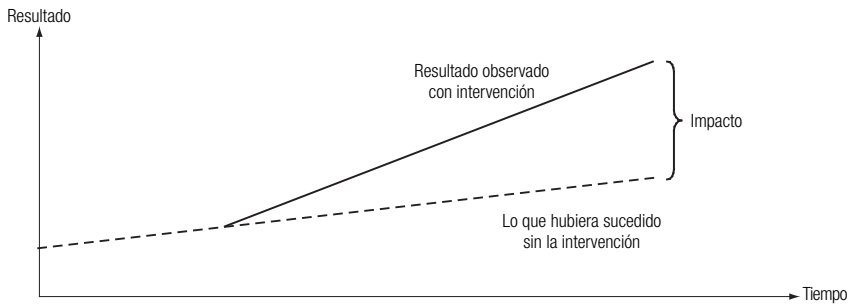
Métodos para valorar el impacto

Las valoraciones de impacto son emprendidas para detectar si en realidad un programa produce los resultados esperados. Para demostrar que una intervención en particular logró el resultado especificado, se requiere reunir ciertas condiciones.⁵ Primero, los cambios generados por la intervención deben mostrar que producen el efecto buscado (puesto de otra manera, el resultado debe responder a la intervención). Segundo, las explicaciones alternativas plausibles para el resultado deben ser contrastadas (hipótesis rivales deben ser rechazadas). Tercero, el mecanismo por el cual se produjo el resultado debe ser explicado; en otras palabras, debe articularse una teoría ligando la intervención al resultado. Finalmente, debe ser posible replicar el resultado en condiciones similares. Con la investigación apropiada, algunas correlaciones pueden ser trasladadas a explicaciones causales creíbles.

A este respecto, el principio de la valoración de impacto fundamental es la necesidad de comparar la situación observada con intervención con lo que hubiera sido de no haber habido intervención alguna. La diferencia en los resultados obtenidos entre estos dos estados constituye el impacto de la intervención como se ilustra en la Figura 10.4.⁶ Mientras que la situación contrafactual no puede ser observada o conocida con certeza completa, el concepto de comparar los resultados observados con este estado hipotético es la base de todos los

acercamientos válidos para determinar impactos. Las comparaciones válidas implican que el efecto neto de la intervención está aislado de todo otro factor externo o confuso que influye en los resultados definidos. Por ejemplo, los esfuerzos para mejorar el desempeño de las empresas mediante vales por servicios de consultoría que pudieron haber sido llevados a cabo durante un periodo de rápida expansión económica ayudada por deducciones de impuestos sustanciales, reformas regulatorias agresivas y una explosión de la demanda por parte de los consumidores. Dadas estas condiciones, se puede asumir que las empresas participantes hayan obtenido un crecimiento significativo incluso la en ausencia del programa de vales. Como resultado, la pregunta central no es si las empresas participantes crecieron, sino más bien si las mismas empresas crecieron más de lo que se hubiera esperado en caso de no haber participado en el programa de vales. Así, el mayor reto en la valoración del impacto es estimar el efecto de los programas después de eliminar factores externos que afectan los resultados. Estos factores pueden incluir eventos específicos o tendencias de largo plazo en industrias, regiones o países en particular como en el ejemplo citado antes. Estos factores también pueden incluir actividades de desarrollo que se estén llevando a cabo dentro de las PyMEs participantes.⁷

Figura 10.4 **Impacto de una intervención**



De igual manera, dentro de lo posible, la valoración del impacto requiere tomar en cuenta el aspecto voluntario del programa. Las PyMEs toman parte en los programas por decisión propia. Algunos integrantes de la población objetivo pueden estar más inclinados a participar debido al mayor interés, motivación u otra condición dentro de la empresa. Este proceso de auto-selección puede sesgar los resultados si los factores que conducen a las compañías a participar están relacionados con resultados específicos bajo estudio.⁸ Por ejemplo, las iniciativas que se enfocan a proveer mayor acceso a financiamiento de largo plazo para la compra de activos fijos posiblemente atraigan compañías en crecimiento con una administración progresista que reconozca el potencial de las oportunidades

de mercado, y que estén dispuestas a asumir riesgos determinados con la esperanza de cosechar retorno financiero, y tengan garantía suficiente para respaldar el préstamo. Es probable que las mismas características estén asociadas con el crecimiento futuro de las ventas. Sería inapropiado comparar este segmento de la población de empresas con otras PyMEs que estén luchando por sobrevivir. Al hacerlo se corre el riesgo de sobreestimar el impacto del programa de asistencia financiera. Como se argumenta adelante, se debe tener cuidado al tomar en cuenta el potencial de sesgo de selección al estimar el impacto de los programas de asistencia a negocios.

Aunque hay numerosas variaciones, el menú de opciones disponibles para valorar las iniciativas dirigidas a las PyMEs se limita a cuatro métodos básicos sustentados en el tipo de controles usados para aislar el efecto de los programas de otros factores desconcertantes: experimentos con muestreo aleatorio, cuasi-experimentos con controles ya establecidos, juicio de participantes y opinión experta, y técnicas no experimentales con controles reflexivos.⁹ La fuerza de la inferencia causal que puede obtenerse del análisis depende de que tan bien el enfoque particular utilizado puede enfrentar la prueba de la validez.¹⁰

Sin importar el propósito o diseño de la iniciativa, toda valoración de impacto requiere emplear uno o más de los métodos siguientes:¹¹

1. *Experimentos con muestreo aleatorio.* La regla de oro en valoración de impacto es el diseño experimental con muestreo aleatorio de los grupos tratados y de control. En este enfoque, las PyMEs en el grupo de tratamiento reciben asistencia, aquellos en el grupo de control reciben un tipo de asistencia alternativo o ninguna. El elemento crítico de este diseño es la aleatoriedad. Lo aleatorio en este caso no significa al azar; se requiere cuidar que cada compañía tiene igual oportunidad de ser seleccionada para cualquier grupo. El muestreo aleatorio ayuda a garantizar que los dos grupos son similares en todo y que cualquier factor externo que influya en los resultados está presente en ambos grupos. Por ejemplo, el muestreo aleatorio ayuda a asegurar que ambos grupos de PyMEs sean similares en términos de la proporción de empresas que intrínsecamente son más receptivas a efectuar los cambios requeridos en las prácticas de negocio, o en que las fluctuaciones en las condiciones del mercado afecten de igual forma a ambos grupos. De esta manera, el grupo de control sirve como uno en una situación contrafactual ideal. Debido a esta capacidad de comparación, las aseveraciones en el sentido de que las diferencias observadas en el resultado entre dos grupos son resultado directo del programa son más difíciles de refutar.

Las evaluaciones que utilizan el diseño experimental son bastante comunes en las áreas de salud, seguridad social y educación para probar la eficacia de nuevos enfoques (ver Recuadro 10.1). Sin embargo, aunque este enfoque es

muy fuerte, no ha sido utilizado de forma extensa en la evaluación del impacto de programas de apoyo a negocios. Hay varias razones para esto. Primero, en ocasiones, las consideraciones políticas hacen difícil asignar PyMEs a diferentes grupos: políticos y administradores de programas tienen ya renuencia a proveer diferentes servicios o a negar el servicio a las compañías asignadas de manera aleatoria al grupo de control. Segundo, frecuentemente difícil mantener las condiciones del experimento: aunque las PyMEs sean estadísticamente equivalentes al inicio del programa, algunos participantes pueden rehusarse a participar o pueden salirse del programa. Además, los servicios prestados a las PyMEs pueden no estar estandarizados y pueden cambiar en el tiempo mientras el programa evoluciona. Finalmente, las evaluaciones utilizando un diseño experimental tienden a ser costosas y difíciles de administrar.

Recuadro 10.1 Ejemplos de diseños experimentales

Experimento argentino de transición de seguro de desempleo a empleo.* El programa Proempleo proveía un subsidio salarial y capacitación especializada como medio de ayudar a la transición del seguro de desempleo a un trabajo regular. Los participantes fueron localizados en dos municipios adyacentes y fueron registrados en programas de seguro de empleo. Los participantes del seguro de desempleo (958 hogares) fueron asignados de manera aleatoria a uno de tres grupos aproximadamente del mismo tamaño: a) aquellos a los que se les entregó un vale que le daba derecho al empleador a recibir un considerable subsidio salarial, b) aquellos que recibieron capacitación voluntaria en habilidades junto con el vale, y c) aquellos que no recibían servicio alguno y fungían como grupo de control.

La evaluación intentó medir el impacto directo del experimento sobre empleo e ingreso de aquellos que recibieron el vale y la capacitación. Se aplicó una encuesta de referencia y varias encuestas de seguimiento durante 18 meses. Se emplearon métodos de doble-diferencia y de variables instrumentales para hacer frente a potenciales sesgos experimentales, incluyendo conformidad selectiva. En comparación con el grupo de control, los receptores tuvieron una probabilidad significativa mucho más alta de emplearse después de los 18 meses, aunque sus salarios actuales no eran mayores. El impacto estuvo confinado en su mayor parte a mujeres y trabajadores jóvenes.

* Galasso, Ravallion y Salvia (2001).

2. *Cuasi-experimentos con controles definidos.* En situaciones donde el diseño experimental no es factible o es impráctico, el enfoque siguiente y más elegido es un diseño cuasi-experimental. Como en el diseño anterior, el cambio en el

desempeño de las PyMEs participantes se compara al de otras PyMEs similares que no han recibido apoyo. Sin embargo, en este caso, la asignación de los dos grupos no es aleatoria. Más bien, se construye un grupo de comparación después del hecho. En la medida que los dos grupos son similares, las diferencias observadas pueden ser atribuidas al programa con mayor grado de confianza. Las comparaciones válidas requieren que los dos grupos sean similares en términos de su composición respecto a la característica clave, exposición a eventos y tendencias externas, y a la propensión para participar en el programa.¹²

Existen varios tipos de diseños que caen en esta categoría general. Estos se discuten a continuación en el orden de su habilidad para hacer frente a factores de confusión:

- *Regresión discontinua.* En esta propuesta, los resultados de una medición en particular son usados para asignar las metas a los grupos de intervención y control de una manera explícita y consistente. La diferencia en el desempeño post-implementación entre ambos grupos es comparada, controlando estadísticamente las variables usadas en el proceso de selección. Por ejemplo, los resultados en relación con el cumplimiento de crédito de las PyMEs pueden ser usado para calificar firmas en la participación en un programa de préstamos; un caso de selección administrativa. Asumiendo que se utilice un límite definido para determinar la elegibilidad, el efecto neto del programa puede estimarse después de ajustarse por la variable de selección original.
- *Controles equilibrados estadísticamente.* Esta propuesta utiliza técnicas para asegurar que tanto la intervención como los controles sean tan equivalentes como sea posible respecto de las características del resultado. En general, esto implica utilizar regresión multivariada en la cual la influencia del programa se estima después de controlar otras variables que pueden afectar los resultados. Por ejemplo, el modelo estadístico usado para estimar el efecto de un programa de consultorías en la productividad empresarial puede incluir diversas variables de control como el tamaño de la empresa, su ramo industrial, localización geográfica, su régimen de constitución y el capital accionario inicial, así como factores que influyeran en la selección. La selección es tratada por medio del uso de regresiones en dos etapas u otras técnicas involucrando variables instrumentales.¹³ En la propuesta de dos etapas, una ecuación inicial es usada para modelar el proceso de selección. El resultado de este análisis (razón inversa de Mills), entonces, es incorporada en una segunda ecuación junto con otras variables de control para estimar los resultados. Como tal, esta propuesta toma en cuenta explícitamente el sesgo potencial de selección.
- *Controles emparejados.* Una propuesta algo menos sofisticada involucra un grupo de comparación que se asemeje al grupo tratado tanto como sea posible basándose en las características que se hayan considerado importantes para

explicar los resultados. Por ejemplo, las compañías pueden ser emparejadas estadísticamente con base en el mismo número de variables descritas en las técnicas previas. Las diferencias en el desempeño post-intervención entre los dos grupos son calculadas sin mayor ajuste estadístico. Sin embargo, puede ser difícil encontrar contrapartes para participantes que simultáneamente se apoyen en todos los criterios, *v. gr.* otra compañía del mismo tamaño, industria, localización geográfica, régimen de constitución, etcétera.

- **Controles genéricos.** La última propuesta son los indicadores de desempeño para la población de la cual se extrajeron los objetivos del control. Por ejemplo, el crecimiento anual en ventas entre las compañías participantes puede ser comparado contra promedios de la industria, y donde cualquier diferencia resultante es atribuida al programa. Sin embargo, los controles genéricos pueden ser incapaces de asegurar que puedan ser comparadas con participantes y deben ser usados con precaución.

A pesar de su complejidad, los diseños cuasi-experimentales han sido usados para evaluar un amplio espectro de programas de apoyo para el desarrollo. Se muestran ejemplos en el Recuadro 10.2.

Recuadro 10.2 Ejemplos de diseños cuasi-experimentales

Programa del Centro de Recursos Industriales (IRC, por sus siglas en inglés).¹ El programa fue establecido en 1988 para ayudar a pequeños y medianos productores con la actualización de sus prácticas de negocio y la modernización de sus capacidades de producción con el propósito de estimular el crecimiento económico en Pensilvania. Los IRCs están diseñados con el propósito de alcanzar esto a través de un conjunto de actividades extensivas que involucran una combinación de servicios de consultoría y entrenamiento. Desde su inicio, el gobierno estatal ha comprometido aproximadamente \$84 millones al programa.

Una evaluación extensiva del programa fue llevada a cabo en 1999. La evaluación incluía una valoración del impacto de los servicios en las compañías participantes en relación con el crecimiento en productividad y la producción. Para ayudar a controlar el sesgo potencial de selección, el análisis utilizó un procedimiento en dos etapas.² El primer paso implica estimar la probabilidad de las compañías de volverse clientes de la IRC en función de las características de la empresa. El segundo paso implica estimar el impacto del programa IRC en las compañías después de controlar los factores que afectan el crecimiento de la productividad y la producción, así como del sesgo potencial de selección. El modelo aproximado se apoya en una función de producción Cobb-Douglas modificada que incluye factores específicos de las plantas.

El análisis está basado en un panel de datos para plantas manufactureras individuales —el Arreglo de Datos Longitudinales para Investigación (CDLI)— actualizada por el Centro de Estudios Económicos en la Oficina de Censos de

Recuadro 10.2 Ejemplos de diseños cuasi-experimentales (cont.)

EU. La preparación de datos provee identificadores detallados en el ámbito de la planta sobre embarques, empleos, costo de los factores, industria y otros aspectos legales y administrativos. Se compila de los Censos de Manufactureros llevado a acabo cada cinco años y la Encuesta Anual de Manufactureros. El CDLI fue usado para obtener datos tanto de clientes como de quienes no lo son en Pensilvania. Las compañías incluidas en la base de datos administrativa del IRC fueron relacionadas con el CDLI usando procesos de emparejamiento desarrollados por el Centro de Estudios Económicos. El proceso de emparejamiento identificó 2839 establecimientos únicos de clientes del IRC en los años de los censos con base en su número permanente de planta (PPN, por su siglas en inglés). El grupo de comparación incluyó una cantidad similar de compañías.

El análisis demostró que el programa tuvo un impacto significativo en los clientes del IRC. Para la cohorte pre-92, la diferencia en la tasa crecimiento en producción y productividad directamente atribuible a los servicios del IRC se estimó al 1.8 por ciento y 3.6 por ciento por año en un periodo de diez años.

Programa de Investigación para la Innovación en los Pequeños Negocios (SBIR, por sus siglas en inglés).³ El programa SBIR fue establecido por el Congreso de los EU. en 1982. La legislatura que lo autorizó mandó que todas las agencias federales que gastaran más de \$100 millones anualmente en investigación externa apartaran un porcentaje de esos fondos para premios a los pequeños negocios. Con el tiempo, el porcentaje se ha incrementado al 2.5 por ciento. Entre 1983 y 1995, las pequeñas empresas han recibido más de \$6 mil millones bajo este programa.

Una evaluación del programa fue llevada a cabo en 1996 enfocándose en el impacto del fondo SBIR sobre el crecimiento de las ventas y el empleo. El análisis involucró la comparación del desempeño de las firmas que recibieron premios de SBIR con compañías similares que no participaron en el programa, controlando la edad de la empresa, localización geográfica, financiamiento previo de la empresa y actividad del capital de riesgo en la región e industria en general. La muestra se compuso de 541 firmas que recibieron el premio SBIR Fase II en los primeros tres ciclos del programa y 594 firmas para del grupo de emparejamiento.

Los datos fueron reunidos desde registros del programa e información contenida en directorios y bases de datos públicamente disponibles.

El análisis demostró que las empresas premiadas por SBIR experimentaron mayor crecimiento tanto en ventas como en empleo que otras empresas similares, pero estos efectos estuvieron confinados a áreas que atraían significativamente al capital de riesgo.

1. Nexus Associates, Inc. (1999).

2. Véase las publicaciones siguientes para mayor información sobre el procedimiento usado en este análisis: Maddala (1994), Jarmin (1997), y Heckman (1974 y 1979).

3. Lerner, Josh (1996).

- *Juicio de los participantes y opinión experta.* El enfoque final se apoya en personas que están familiarizadas con la intervención para hacer juicios respecto a su impacto. Esto puede involucrar a participantes del programa o expertos independientes. En cualquier caso, se solicita a los individuos que estimen en que medida el desempeño mejoró como resultado del programa; en efecto, para comparar su desempeño actual con lo que hubiera sido en ausencia del programa.

Aunque este enfoque es bastante común, el mismo está cargado de problemas. Requiere que las personas sean capaces de determinar el efecto neto de la intervención basadas únicamente en su propio conocimiento sin referencia a comparaciones explícitas. Sin embargo, puede ser la única opción disponible dadas las restricciones de datos y presupuestales. Cuando se use, se deberá tener cuidado de que las personas consideren la situación contrafactual en sus valoraciones del impacto (ver Recuadro 10.3).

Recuadro 10.3 **Ejemplos de juicio de los participantes y opinión experta**

Subsidio Regional a Empresas.¹ El programa de subsidios comenzó en 1988 por el Departamento de Industria y Comercio en Reino Unido. El programa tiene dos componentes: una tasa uniforme de subsidio que contribuye con el 15% (hasta un máximo de £15000) para la compra de activos fijos, y un subsidio que cubre el 50% de los costos de proyectos idóneos (hasta un máximo de £25000) para desarrollar nuevos productos o procesos dirigidos a la producción comercial. Las empresas con menos de 25 empleados son elegibles para participar.

Una evaluación del programa fue realizada en 1991. Esta se basó en los resultados de una encuesta efectuada sobre una muestra representativa de 100 empresas entre la población de compañías que habían remitido su aplicación después del 31 de diciembre de 1998, y completado sus proyectos. Todas las encuestas fueron realizadas en persona. La encuesta solicitaba a quienes respondían: indicar el impacto de los subsidios en términos de incremento en el retorno, el valor agregado y el empleo. La situación contrafactual fue obtenida a través de sondeos a profundidad sobre lo que las empresas hubieran hecho si no hubiera existido el subsidio y como hubieran financiado sus proyectos en la ausencia de los subsidios.

El estudio sugirió que el impacto de los subsidios para la adquisición de activos fijos era bastante débil, mientras que los subsidios para la innovación habían desempeñado un papel fructífero entre las empresas con potencial genuino de crecimiento.

Extensión de la Sociedad para la Manufactura. En 1988, el Congreso de EU. instruyó al Instituto Nacional de Indicadores y Tecnología (NIST, por sus

Recuadro 10.3 Ejemplos de juicio de los participantes y opinión experta (cont.)

siglas en inglés) establecer un programa para ayudar a las pequeñas manufactureras —establecimientos de manufactura con 500 o menos trabajadores— para mejorar su desempeño competitivo. Desde sus inicios con sólo tres centros, la Extensión de la Sociedad para la Manufactura (MEP, por sus siglas en inglés) se ha expandido hasta 68 centros con más de 400 oficinas en todo Estados Unidos.

Las compañías que recibieron servicios de un centro NIST MEP son encuestadas por una compañía independiente de encuestas un año después de completar algún proyecto. Los proyectos principales son definidos como actividades que requieren ocho o más horas de tiempo del personal del centro o de un proveedor de servicios a terceros. Cada encuesta trimestral, se pregunta a los clientes sobre los impactos experimentados en los doce meses anteriores como resultado directo de los servicios provistos por el centro. Específicamente, se les solicita estimar los cambios en el desempeño resultantes de la participación en el programa. Los resultados de las encuestas sugieren que los impactos agregados respecto a los incrementos en ventas y empleo son bastante grandes en proporción a la inversión federal en el programa: sin embargo, una pequeña porción de las firmas que respondieron a la encuesta soportan el conjunto de los impactos agregados.

Además de las técnicas cuantitativas, NIST MEP ha apoyado investigaciones mediante la técnica de casos enfocadas en compromisos ejemplares fundamentados en modelos conceptuales que relacionan los servicios con resultados del programa.²

1. Segal Quince Wicksteed, Ltd. (1991).

2. Departamento de Comercio de EU (1997).

- *Técnicas no experimentales con controles reflexivos.* Por lo general, las comparaciones antes y después son inválidas pues fallan en controlar otros factores que pudieron contribuir con los resultados observados. Como tales, los resultados de estudios basados exclusivamente en controles reflexivos deben ser tratados con especial cuidado. Dicho lo anterior, este enfoque puede ser válido cuando existe una clara y cercana relación entre el programa y los resultados de interés (véase Recuadro 10.4). Adicionalmente, los controles reflexivos son usados en ocasiones cuando es imposible construir un grupo de control, como en el caso de programas de cobertura total que afectan a todas las compañías en la población objetivo.

En estas cuatro propuestas, es posible usar los datos del programa para mejorar el análisis. Es frecuente que los programas no sean administrados de manera uniforme —esto es, la intervención puede variar en intensidad entre

los miembros de la población objetivo. Por ejemplo, mientras que algunas PyMEs pueden recibir 40 horas de asistencia técnica bajo un esquema para proporcionar servicios de consultoría con costo compartido, otras pueden recibir significativamente mayor o menor apoyo. El impacto de diferentes niveles de intensidad (algunas veces llamado el efecto de dosificación) puede ser estudiado bajo las cuatro propuestas. Dependiendo del grado de variación, esta propuesta puede fortalecer las inferencias causales que pueden ser obtenidas del análisis.¹⁴

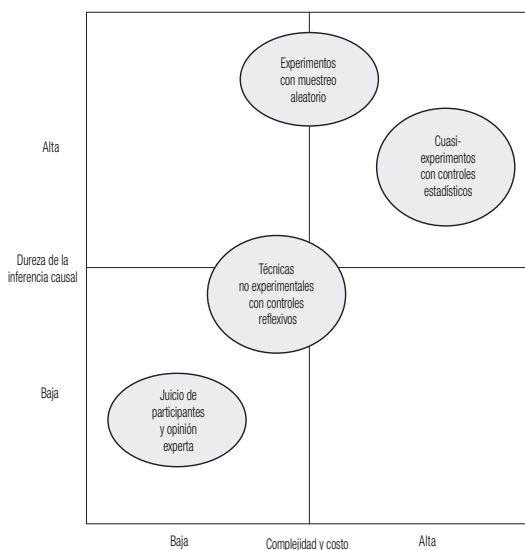
Recuadro 10.4 **Ejemplo de técnica no experimental con controles reflexivos**

COMPITE. En el programa COMPITE de México, consultores certificados trabajan con las compañías en un periodo de cuatro días para mejorar el desempeño de una línea de producción en cuatro dimensiones —productividad, tiempo de carga en manufactura, inventario de productos en proceso y requerimientos de espacio en planta. El primer día se invierte entrenando hasta 20 empleados en técnicas de producción esbeltas. El segundo día se dedica a medir el desempeño actual, diagnosticar problemas e idear medios para mejorar el funcionamiento. El tercer día se implementan los cambios. Otro conjunto de indicadores son tomadas el cuarto día, y comparadas con aquellas tomadas antes de implementar los cambios. Aunque éste es un diseño antes-después, debido a la estrecha relación de la intervención del programa hay poca duda de que el programa fue el origen del cambio en los resultados observados. Sin embargo, el diseño antes-después no sería particularmente duro, si el asunto se turna hacia el impacto de COMPITE en el ámbito general de los resultados de la empresa uno o dos años después de que la compañía participó en el programa.

Elección de un método apropiado

Las cuatro propuestas para valorar los impactos pueden ser aplicadas a una gran variedad de preguntas que puede plantear el gobierno en relación con los programas de apoyo a negocios. La elección de la propuesta a ser usada para la valoración de un impacto en particular requiere tomar en cuenta varios factores. Cada propuesta tiene fortalezas y debilidades como se ilustra en la Figura 10.5. Por ejemplo, los diseños experimentales proveen evidencia concluyente de causalidad, pero pueden ser difíciles y costosos de administrar. Los diseños no experimentales son por lo general más fáciles de implementar, pero puede que no ofrezcan inferencias causales suficientemente concluyentes.

Figura 10.5 Intercambios de los diseños de evaluación



La figura sugiere que los enfoques más sofisticados como son los diseños experimentales y cuasi-experimentales deben ser usados donde garanticen la solidez de las inferencias causales que pueden obtenerse. Sin embargo, la solidez adicional viene con un mayor costo. Por lo tanto, estas propuestas sólo serán apropiadas cuando se estén considerando inversiones más cuantiosas. Por ejemplo, numerosas iniciativas son establecidas inicialmente como programas piloto con la expresa intención de expandir o duplicar la iniciativa si tiene éxito. Dependiendo de la magnitud de la inversión requerida, puede tener sentido para los gobiernos comprometer recursos en evaluaciones rigurosas para ayudar a tomar una decisión informada. Sin embargo, la habilidad de usar estas propuestas es contingente en poder obtener los datos requeridos y en contar con el tiempo suficiente para diseñar e implementar el estudio.

La regla general es que debe usarse el mejor diseño posible desde una perspectiva metodológica, tomando en cuenta lo significativo de la inversión, así como consideraciones prácticas relacionadas con la factibilidad técnica.

Obtención de los datos

Independientemente del enfoque, toda valoración de impacto requiere datos exactos y confiables. Como mínimo, es necesario obtener datos de indicadores de los resultados de entidades que fueron afectadas por la iniciativa. Si se tienen considerados análisis comparativos, los datos sobre los indicadores de los resultados y otras variables será necesario para miembros de los grupos de tratamiento y de

control.^{15,16} En todos los casos, se necesitan datos de referencia (preferentemente indicadores múltiples) junto con datos después de la intervención.

Existen sólo dos opciones para obtener los datos requeridos:

1. *Obtener los datos existentes en el gobierno y otras organizaciones.* Debido al costo de recopilar los datos es preferible aprovechar, dentro de lo posible, los datos existentes. Las agencias gubernamentales de estadística en muchos países llevan a cabo encuestas de empresas y hogares de forma rutinaria que pueden ser utilizadas en la valoración de impacto. Entre éstas se tienen encuestas de ingreso-gasto nacionales, encuestas de ingreso-gasto de los hogares, encuestas del mercado laboral, y varias encuestas industriales. Otras organizaciones como bancos, uniones de crédito y cooperativas mantienen datos sobre un gran número de compañías como parte de sus operaciones actuales. Deben explorarse estas fuentes para determinar si están disponibles los datos requeridos para la valoración de impacto.
2. *Comisionar encuestas especiales.* Sin embargo, en muchos casos, el único recurso es llevar a cabo encuestas emprendidas específicamente para la valoración del impacto. Existen varios aspectos críticos que requieren atención para diseñar y administrar exitosamente la encuesta. El tipo de encuesta escogida, la redacción de las preguntas, las estrategias de muestreo, el seguimiento y los procedimientos de captura, todos son importantes para orientar la exactitud y utilidad de los resultados de la encuesta. Es pertinente hacer énfasis especial en el muestreo; es esencial utilizar muestreo de probabilidad para asegurar la validez de los resultados, la estratificación debe ser considerada para mayor eficiencia y para asegurar que la muestra representa apropiadamente a la población objetivo en conjunto.¹⁷

Además, las cuatro propuestas requieren registros administrativos para identificar y caracterizar a los receptores del servicio, así como la naturaleza del servicio recibido. Aún más, ciertas técnicas requieren información adicional. Por ejemplo, para poder utilizar la regresión discontinua, el programa debe llevar datos sobre las variables usadas para determinar la elegibilidad o calificación del participante.

Recomendaciones

Los principios y las técnicas básicas para llevar a cabo una evaluación están bien establecidos. Sin embargo, sólo de manera reciente han sido aplicados a los programas de apoyos a negocios. Esta experiencia apunta hacia varias recomendaciones importantes:

- *Clarifique los objetivos, las metas y la lógica del programa subyacente.* La valoración de impacto requiere una articulación cuidadosa de los objetivos de las iniciativas,

los cambios específicos que se espera sean obtenidos como resultado de las iniciativas, y la relación causal entre actividades definidas y los resultados esperados. Esto debe ser resumido en un modelo lógico de programa formal o marco de registro. Este ejercicio se realiza mejor como parte del proceso de diseñar nuevas iniciativas, en lugar de como una actividad inicial de una evaluación *ex post*.

- *Planeé la evaluación al inicio del programa.* La valoración de impacto debe ser planeada tan temprano como sea posible, preferentemente antes de que la iniciativa haya sido lanzada. Éste es claramente el caso de las propuestas experimentales con muestreo aleatorio; con todo, en general los diseños *ex ante* tienden a ser más fuertes dado que las medidas para obtener los datos requeridos pueden estar disponibles antes de la implementación del programa. Para este propósito, se debe exigir que todos los programas desarrollen un plan formal de evaluación como condición para recibir financiamiento. El plan debe describir el propósito de la evaluación, preguntas específicas que deberán hacerse, el diseño de la evaluación, el proceso y medidas de resultados, estrategias de recopilación de datos, análisis posibles, informes y otros métodos para comunicar los resultados, itinerario, papel y responsabilidades del personal y contratistas externos, si los hay, y un presupuesto estimado.¹⁸
- *Establezca una línea de base y programe los registros.* Los programas deben recopilar información de base sobre las características y desempeño de los objetivos del programa. Incluso, todos los programas requieren llevar registros completos y exactos como parte de la implementación del programa, incluyendo la naturaleza y magnitud de los recursos comprometidos a compañías o instituciones en particular.
- *Reconozca que la valoración de impacto es explícitamente para demostrar causalidad.* Mientras que algunas personas dentro de la comunidad donante (desarrollo de cooperación financiera en países en vía de desarrollo) están exigiendo mayor responsabilidad en términos de efectividad, otras se quejan sobre la futilidad de tratar de establecer dicha atribución. Es difícil reconciliar estas posiciones. La valoración de impacto está dirigida específicamente a la demostración de las iniciativas en particular producen los resultados deseados —dicho de otra forma, con el propósito de establecer causalidad. Todas las evaluaciones deben estar diseñadas para mostrar el efecto, eliminar explicaciones alternativas y explicar mecanismos causales. Replicar de los resultados en situaciones similares puede añadir mayor credibilidad.
- *Incluya comparaciones válidas en el análisis.* Valorar el impacto de las iniciativas dirigidas a las PyMEs involucra comparar fenómenos observados vs. la situación contrafactual —una situación hipotética que hubiera ocurrido en ausencia del programa. Se necesitan el muestreo aleatorio, los controles ya establecidos o

los controles reflexivos para aislar el impacto del programa de otros factores que afectan los resultados.

- *Utilice múltiples métodos.* Los acercamientos descritos en esta investigación no son mutuamente excluyentes. Cuando sea posible, deben utilizarse múltiples técnicas para valorar el impacto de una iniciativa específica. Resultados similares desde diferentes métodos pueden agregar a la credibilidad de los hallazgos. Además, debe utilizarse investigación cualitativa para complementar las técnicas cuantitativas, para proporcionar intuiciones hacia los mecanismos causales específicos que entran en juego para generar los resultados.
- *Reconozca que suficientemente bueno es suficientemente bueno.* Los gobiernos y otros inversionistas deben esforzarse por ser lo más rigurosos posible. Sin embargo, deben tomarse en cuenta consideraciones prácticas al diseñar e implementar una valoración de impacto. Las restricciones en datos, tiempo y presupuesto pueden hacer inviable la adopción de ciertas propuestas. Los inversionistas deben aceptar estas limitaciones.
- *Comprometa recursos.* El monto de dinero dedicado a la evaluación queda a discreción del gobierno. El mismo debe estar dispuesto a comprometer el nivel de recursos necesario para diseñar evaluaciones fuertes, recopilar datos vitales, llevar a cabo los análisis requeridos y reportar los resultados.

Notas

1. Las definiciones varían entre países y usualmente se basan en el número de empleados, ventas anuales o activos. De manera típica, las microempresas son definidas como empresas con hasta 10 empleados, las pequeñas empresas tienen entre 10 y 50-100 empleados y la empresa mediana entre 50 a 100-250 empleados.
2. La economía enseña que la competencia perfecta lleva a la asignación Pareto-eficiente de los recursos siempre que se den ciertos supuestos. Requiere que exista un gran número de productores y consumidores en un mercado dado, ninguno de los cuales debe poder influir en los precios por sí mismo; los actores económicos son racionales; todos los recursos (incluyendo la información) tienen movilidad perfecta; y los costos de transacción son cero. Desafortunadamente, los mercados en el mundo real nunca concuerdan con este ideal. Además, la eficiencia de Pareto no garantiza un resultado equitativo. Tres tipos de imperfecciones de mercado son relevantes para los programas de apoyo a las empresas. La información asimétrica puede llevar a pasar por alto el apoyo o inversión requeridos. Adicionalmente, puede que las compañías no obtengan todos los beneficios relacionados con una acción en particular, llevando a una inversión por debajo de lo esperado desde una perspectiva social. Estas externalidades surgen cuando la producción (o el consumo) de un bien afecta a otros aparte de los directamente involucrados en la transacción. Finalmente, en algunos casos, el apoyo a las empresas puede ser un bien público. Los bienes públicos tienen dos propiedades únicas: primero, el consumo de un bien público por un consumidor no afecta la habilidad de otro consumidor para beneficiarse de él (*no rivalidad*), y segundo, es difícil impedir a las personas beneficiarse del bien, incluso en el caso en que no estén dispuestos a pagar por él (*sin exclusión*). Por estas dos características, los bienes públicos tienden a suministrarse por debajo de lo requerido por la economía de mercado.

3. De manera más formal, las evaluaciones deben ser diseñadas para probar la hipótesis nula en la cual el crédito fiscal para I&D no conduce a inversión adicional. Técnicas cualitativas o cuantitativas son usadas para determinar si la hipótesis nula puede ser rechazada con un nivel de confianza razonable.
4. Preferentemente, el modelo lógico del programa debe ser desarrollado durante la fase de diseño del proyecto.
5. Véase Mosteller y Tukey (1977) para una discusión de las condiciones requeridas para demostrar causalidad.
6. En ocasiones, a esto se le refiere como “adicionalidad”.
7. Las amenazas a la validez interna son generalmente agrupadas en varias categorías generales, como son: eventos externos, arrastre seglar, maduración, regresión y agotamiento. Los lectores interesados en explorar estos conceptos de manera más detallada pueden referirse a Cook y Campbell (1979).
8. Un tipo similar de sesgo de selección puede ocurrir cuando las organizaciones seleccionan a los participantes basados en características determinadas (a esto se refiere en la bibliografía como selección administrativa).
9. Una quinta propuesta, el estudio estructurado de casos, puede ser usada para estudiar el impacto de la intervención en participantes y puede ser utilizada para complementar técnicas cuantitativas. A diferencia de las otras técnicas descritas en la investigación, el estudio de casos se asienta en extensas descripciones narrativas y otras evidencias para afirmar que la intervención causó el resultado observado. En general, el estudio de casos involucra múltiples fuentes de información, incluyendo: observación directa, entrevistas, documentos y mecanismos físicos. En todos los casos, los modelos lógicos de los programas desempeñan un papel crítico. Mientras que el estudio de casos puede proveer abundantes explicaciones de cómo y porqué el programa tuvo efecto en ciertas empresas, es difícil generalizar los resultados más allá de las empresas estudiadas. Esto es particularmente cierto para programas con clientes y servicios variados.
10. Es importante notar que estas propuestas no son mutuamente excluyentes; pueden ser combinadas bajo ciertas circunstancias para mejorar el análisis.
11. Algunos de los métodos para valorar el impacto de los programas de apoyo a negocios son bastante complejos, requieren antecedentes en estadística y econometría. Por lo mismo, una explicación detallada de los aspectos técnicos involucrados en llevar a cabo cada tipo de estudio está fuera del alcance de esta investigación.
12. La cuestión de la validez de un grupo de comparación es medular para esta propuesta. Idealmente, el grupo no participante debería ser similar al grupo participante respecto a las variables que afectan la medida del resultado, pero no debe haber recibido el apoyo para negocio a través del programa del gobierno. “Similar” en este contexto se refiere a la distribución de valores de estas variables, i. e. la media y el rango.
13. Ver Heckman (1985). Una técnica similar conocida como Parcialidad de Resultados Emparejados (Propensity Score Matching) puede ser usada para controlar por medio de la selección. Véase Jalan y Ravallion (por publicarse).
14. El uso de datos de dosificación es especialmente importante en el caso de programas de cobertura total, donde grupos de comparación pura no están disponibles.
15. Los requerimientos de datos son específicos para las medidas de resultados de interés y la naturaleza del análisis que será realizado. De manera adicional a las medidas de resultados, pueden requerirse datos para las variables explicativas usadas en el análisis de regresión, incluyendo variables instrumentales usadas para controlar el sesgo potencial de selección.

16. Generalmente se requieren grandes muestras para los diseños cuasi-experimentales. Para calcular el tamaño requerido de la muestra es importante considerar tres factores: la posible variación de las medidas de resultados, el grado de confianza requerido y la precisión deseada de los estimados. Lo último puede ser pensado como el mínimo efecto deseado por los inversionistas; cualquier cosa menor no será considerada un éxito de interés particular. Un ejemplo concreto: asuma que los inversionistas están interesados en determinar si el programa ha obtenido como resultado un crecimiento en las ventas. Además, asuma que el programa debe estar dirigido a incrementar el crecimiento 5 puntos porcentuales más de lo que ocurriría en ausencia del programa. Visto de otra forma, si las compañías en el grupo de comparación crecieron en promedio un 10 por ciento por año, los inversionistas desearían ver crecer anualmente a las compañías participantes un promedio del 15 por ciento. Dada una desviación estándar del 30 por ciento, un intervalo de confianza del 95 por ciento y un poder de prueba del 90 por ciento, se deben obtener datos de aproximadamente unas 620 compañías participantes y no participantes.
17. Kish (1965).
18. El plan de evaluación debe ir más allá de los aspectos relacionados con la valoración de impacto y hacerse cargo de otras facetas del desempeño, incluyendo excedentes, eficiencia operativa, autosuficiencia financiera y otros asuntos.

Referencias bibliográficas

- COOK, Thomas D. y Donald T. CAMPBELL (1979), *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- GALASSO, Emanuela, Martin RAVALLION y Agustin SALVIA (2001), "Assisting the Transition from Workfare to Work: Argentina's Proempleo Experiment". World Bank, septiembre 24.
- HECKMAN, J. (1974), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Development Variables and Simple Estimator for such Models", *The Annals of Economic and Social Measurement*, 5:475-492.
- HECKMAN, J. (1979), "Sample Selection Bias as Specification Errors". *Econometrica*, 47:153-161.
- HECKMAN, James y Richard ROBB (1985), "Alternative Methods of Evaluating the Impact of Interventions: An overview". *Journal of Econometrics* 30:239-67.
- KISH, Leslie (1965), *Survey Sampling*. Nueva York: John Wiley and Sons.
- JALAN, Jyotsna y Martin RAVALLION (por publicarse), "Estimating the Benefit Incidence of an Antipoverty Program by Propensity Score Matching", *Journal of Business and Economic Statistics*.
- JARMIN, R.J. (1997), "Evaluating Impact of Manufacturing Extension on Productivity Growth", Manuscrito, Centro de Estudios Económicos, Oficina de Censos de EU., Washington, DC 1997.
- LERNER, Josh (1996), "The Government as Venture Capitalist: The long-run impact of the SBIR Program", investigación en curso 5753, Oficina Nacional de Investigaciones Económicas.
- MADDALA, G.S. (1994), "Limited Dependent and Qualitative Variables in Econometrics", *Econometric Society Monographs* núm. 3, Cambridge University Press. 1994.

MOSTELLER, F.M. y J.W. TUKEY (1977), *Data Analysis and Regression*. Reading, MA: Addison-Wesley.

NEXUS ASSOCIATES, INC. (1999), "The Pennsylvania Industrial Resource Centers: Assessing the record and charting the future".

SEGAL QUINCE WICKSTEED, LTD. (1991), "Evaluating of Regional Enterprise Grants - Second Stage".

DEPARTAMENTO DE COMERCIO DE LOS EE. UU. A. (1997), *MEP Successes: A case study approach*. Gaithersburg, MD: NIST Special Publication 916.

Capítulo 11

Evaluación de los programas de capacitación: impacto en el ámbito local

por
Randall W. Eberts
y
Christopher J. O'Leary
Instituto W. E. Upjohn para la Investigación del Empleo,
Kalamazoo, MI, Estados Unidos

Introducción

Las evaluaciones son parte integral de muchos de los programas públicos de capacitación para el trabajo en los Estados Unidos. Desde los años sesenta se han dedicado importantes recursos y tiempo para tratar de entender el impacto de los programas de empleo y capacitación. De hecho, muchos avances en las técnicas para evaluar programas sociales fueron desarrollados durante los estudios sobre los efectos de los programas de empleo (Stromsdorfer y Farkas, 1980). Los diseñadores de las políticas llegaron a apreciar el valor de evaluaciones objetivas y rigurosas para apoyar sus decisiones respecto a estos programas. De hecho, la legislatura que autorizó la Job Training Partnership Act (JTPA) [Decreto de Asociación para la Capacitación para el Trabajo] de 1982 asignó por mandato un sistema de medición del desempeño durante la ejecución y una evaluación neta de impacto. La última fue hecha usando el método que se convirtió en la regla de oro de la evaluación —un experimento de campo que involucra muestreo aleatorio (Orr *et al.*, 1996). Como observaron Kluge y Schmidt (2002), el énfasis americano en evaluar los programas de empleo y capacitación es característicamente diferente de la práctica usual en Europa.

No obstante, incluso en los Estados Unidos el entusiasmo por las evaluaciones no se extiende a todos los programas de capacitación para el trabajo. Mientras que la mayoría de los programas financiados federalmente son evaluados con regularidad, los programas apoyados por los estados han recibido mucho menos escrutinio. Esto es particularmente cierto para aquellos programas de capacitación patrocinados por el estado que proporcionan capacitación personalizada a las empresas como medio para promover el desarrollo económico local y regional. Una razón para esto puede ser la complejidad para evaluar estos programas que comúnmente tienen una multiplicidad de objetivos. La evaluación de los programas con dicho diseño frecuentemente termina en resultados imprecisos y no conclusivos, lo que deteriora la credibilidad y el valor de realizar las evaluaciones. Puede ser no menos importante, la falta de voluntad política y los recursos limitados para llevar a cabo evaluaciones apropiadas en el ámbito estatal y local.

Este artículo revisa evaluaciones previas de programas de capacitación para el trabajo con la mira en las lecciones que pueden ser aplicadas a los programas

estatales y locales. En un sentido real, todas las evaluaciones son locales, pues la participación y los datos son compilados localmente. La evaluación de programas federales de capacitación para el trabajo se enfoca típicamente en el éxito promedio de los participantes del programa, mientras que los programas financiados local y estatalmente con frecuencia buscan tener efecto en todas las condiciones de la economía local. Nosotros contrastamos las evaluaciones de programas federales de capacitación para el trabajo con aquellas de los programas de capacitación financiados estatal y federalmente para resaltar cómo las técnicas de evaluación varían para ajustarse a diferentes características de los programas y del contexto en el que son administradas.

Al evaluar un programa de capacitación para el trabajo, o cualquier programa gubernamental en esa materia, se deben seguir varios pasos básicos, y hemos organizado el artículo alrededor de esos componentes del proceso de evaluación. En primer lugar, se debe entender el propósito del programa de manera que puedan cuantificarse apropiadamente los indicadores de resultados obtenidos y recolectarse correctamente los datos. Segundo, debe entenderse la administración del programa incluyendo el modo en que los servicios son provistos. Tercero, se debe identificar, también, a los grupos que son objetivo para recibir los servicios y las características de aquellos que en realidad participan. Finalmente, la metodología de evaluación apropiada debe ser elegida considerando las restricciones políticas, administrativas y de presupuesto. Adicionalmente, resaltaremos el papel que han desempeñado las evaluaciones en el diseño de las políticas de empleo en los Estados Unidos. También consideramos los factores que impactan en el grado en que son interpretadas y utilizadas las evaluaciones, y ofrecemos sugerencias de cómo las evaluaciones pueden desempeñar un papel importante, no sólo en las decisiones para elaborar las políticas, sino también al ayudar a los administradores locales a mejorar el desempeño de sus operaciones.

Capacitación para el trabajo financiada públicamente en los Estados Unidos

La mayoría de los programas de capacitación para el trabajo financiados públicamente son programas federales. Su propósito primordial es apoyar a los trabajadores para obtener las habilidades necesarias para encontrar un trabajo remunerado. Aunque se reconoce que el desarrollo de la fuerza laboral promueve el desarrollo económico local y regional, los programas federales se enfocan en los beneficios a los individuos y prestan poca atención a sus posibles efectos en las economías locales. Cualquiera de los beneficios que puedan surgir de estos programas en las áreas locales son sólo de interés secundario. Por otro lado, los gobiernos de los estados han tomado un papel más activo al utilizar los programas de capacitación para el trabajo con el propósito de promover el desarrollo

económico local y regional. Durante las pasadas dos décadas, los estados han implementado programas de capacitación que benefician directamente a las empresas. El propósito de estos programas es mejorar el ambiente empresarial dentro del estado brindando capacitación acorde con las necesidades de los negocios locales.

En este artículo, nosotros hacemos la distinción entre los programas federales que benefician a los trabajadores de manera individual y los programas estatales que se dirigen a las grandes empresas con propósitos de desarrollo económico. Las evaluaciones más rigurosas de los programas de capacitación se han enfocado en los programas federales. Si bien estas evaluaciones han sido efectuadas en áreas locales seleccionadas, dichas evaluaciones sólo se enfocan en los resultados de los individuos bajo las formas de empleo y compensación y no se fijan en el efecto de estos cambios en la economía local. Las evaluaciones de los programas estatales, que se enfocan en los esfuerzos para el desarrollo económico, son mucho menos rigurosas; principalmente porque es política y administrativamente más difícil utilizar diseños de muestreo aleatorio cuando se involucran a las empresas y a las jurisdicciones gubernamentales. Por lo tanto, enfocaremos nuestra atención primero a la evaluación de los programas federales y posteriormente regresaremos hacia las formas de evaluar los efectos de los programas de capacitación sobre las economías locales.

Enfoques para brindar capacitación para el trabajo financiada públicamente

Los programas de capacitación para el trabajo financiados públicamente han seguido diversas estrategias para proporcionar sus servicios. Éstos oscilan desde brindar instrucción en habilidades remediales, como son la lectura y la aritmética, hasta ofrecer capacitación en procedimientos minuciosos para desarrollar un nivel más alto en una ocupación específica. La Tabla 11.1 muestra una taxonomía de los diversos tipos de capacitación que prevalecen en los programas gubernamentales. La capacitación para el trabajo es diferente al proceso de educación formal, la cual generalmente se brinda por periodos cortos y se enfoca al dominio de habilidades específicas. Para los programas dirigidos a los individuos, el objetivo de la capacitación para el trabajo consiste en ocuparse del desajuste estructural entre las habilidades de un buscador de trabajo y las necesidades de un empleador, de modo que el individuo pueda volver a trabajar. Este objetivo puede implicar brindar capacitación sobre un tipo específico de equipo o proporcionar capacitación básica acerca del comportamiento adecuado en el lugar de trabajo o acerca de las habilidades para buscar empleo.

La capacitación se realiza en diversos escenarios. El enfoque tradicional utiliza el ámbito de salón de clases, que puede albergar tanto a la enseñanza en

general como la capacitación personalizada. En el otro extremo del espectro está la capacitación en el sitio de trabajo. Este se lleva a cabo en el lugar de trabajo y típicamente en el escenario en el cual el trabajador será asignado eventualmente. Por supuesto, algunos programas de capacitación utilizan una combinación de ambos enfoques, como es el caso de varios programas para jóvenes. La capacitación post-empleo también combina las actividades del aula con prácticas y actividades correspondientes que están directamente vinculadas con la continuidad en el empleo y el ascenso en un empleo específico o en un campo de ocupación.

Tabla 11.1 **Tipos de capacitación para el empleo**

<p>Capacitación para habilidades profesionales Proporcionada en escenario grupal, se denomina como capacitación institucional o de aula y usualmente se brinda para profesiones de demanda general. La capacitación personalizada es diseñada para ajustarse a las peticiones específicas de un empleador con plazas disponibles o empleados ya contratados. Los vales son medios que permiten a los participantes elegir entre temas e instructores aprobados. La capacitación en habilidades proporcionada en un escenario de trabajo experimental es referida como una capacitación en el sitio de trabajo (OJT, por sus siglas en inglés). Cuando el OJT se proporciona a través de una agencia pública, algunas veces, se denomina como experiencia laboral.</p>
<p>Capacitación remedial Es la capacitación que en general busca remediar deficiencias básicas en las habilidades de lectura y matemáticas para preparar a los buscadores de empleo para la capacitación en habilidades.</p>
<p>Capacitación en aula sobre competencias sociales o emocionales (soft skills) Transmite conocimientos sobre el comportamiento apropiado en el lugar de trabajo o acerca de las habilidades para buscar empleo.</p>
<p>Capacitación post-empleo Combina clases en un aula y actividades prácticas con la intención de promover la conservación del empleo y el progreso dentro de una trayectoria profesional determinada.</p>
<p>Programas de capacitación para jóvenes Es la capacitación de habilidades básicas en el contexto del lugar de trabajo. Es un apoyo complementario a la educación en general, para la obtención de credenciales, asesorías especializadas, servicios para transitar de la escuela al trabajo y de la escuela a programas de aprendizaje de determinados oficios, educación intensiva en residencia, y capacitación profesional para el trabajo.</p>

Programas federales recientes de capacitación para el trabajo en los Estados Unidos

La política federal de capacitación para el trabajo tiene sus orígenes en la era de la depresión con los programas de obras públicas del *New Deal* (*Nuevo trato*) en los treinta. Los esfuerzos de capacitación renovados treinta años después fueron influenciados en gran medida por los nuevos objetivos económicos planteados durante la *Lucha contra la Pobreza* del Presidente Johnson. Los programas subsecuentes reflejaron las preferencias políticas hacia diferentes grupos de la población y hacia las realidades económicas de la época. Un resumen de los cuatro principales programas federales de capacitación para el trabajo de la posguerra se proporciona en la Tabla 11.2.

Tabla 11.2 Cronología de los programas federales de capacitación para el trabajo en los Estados Unidos

Programa	Tipos de capacitación	Elegibilidad	Relaciones intergubernamentales
Decreto de Desarrollo de la Mano de Obra y Capacitación (MDTA, por sus siglas en inglés), 1962	Capacitación institucional y en el sitio de trabajo (OJT).	Bajos ingresos y destinatarios de asistencia social.	Financiamiento federal proporcionado directamente desde 12 oficinas regionales a las agencias en las áreas locales. Estructuras similares de administración y reporte.
Decreto Integral de Empleo y Capacitación (CETA, por sus siglas en inglés), 1973	Capacitación en el sitio del empleo, capacitación de habilidades en el aula, capacitación de competencias sociales en el aula, experiencia laboral en agencias públicas y Servicio Público de Empleo (PSE, por sus siglas en inglés).	La capacitación se dirigió a personas de bajos ingresos, destinatarios de asistencia social y jóvenes con desventajas.	Financiamiento federal suministrado a los principales patrocinadores en regiones subestatales, que se contabilizan en cerca de 470. Monitoreo del desempeño con los resultados reportados al Departamento del Empleo de los Estados Unidos (USDOL, por sus siglas en inglés).
Decreto de Asociación para la Capacitación laboral (JTPA, siglas en inglés), 1982	Capacitación en el sitio del trabajo, capacitación de habilidades sociales en el aula, capacitación de competencias sociales en el aula, experiencia laboral en agencias públicas.	Bajos ingresos, destinatarios de asistencia pública, trabajadores desplazados y jóvenes con desventajas.	Financiamiento federal a través de los gobernadores de los estados a concejos industriales privados (PICs, siglas en inglés) en cada una de las 640 áreas de suministro del servicio. Los PICs reportan su desempeño a los gobernadores, quienes reportan al USDOL.
Decreto de Inversión en Mano de Obra (WIA, siglas en inglés), 1998	Capacitación en el sitio del trabajo, capacitación de habilidades personalizadas en las aulas, capacitación de competencias sociales en el aula, y experiencia laboral en agencias públicas.	Acceso a servicios básicos como las habilidades para buscar trabajo y la recomendación para el empleo no está restringido. La capacitación esta dirigida a quienes se les dificulta más volverse a emplear.	Como en la JTPA, pero los PICs fueron menos (600) tableros de inversión para la mano de obra (WIBs, siglas en inglés) con una mayoría de miembros en el sector privado. El monitoreo es reducido de acuerdo con la práctica de la JTPA.

Decreto de desarrollo de la mano de obra y capacitación

Bajo el Decreto de Desarrollo de la Mano de Obra y capacitación (Manpower Development and Training Act, MDTA) de 1962, la capacitación para el trabajo se orientó a las poblaciones de bajos ingresos y destinatarios de asistencia social. Los fondos fueron asignados a las comunidades con base en las estimaciones de la porción de la población por debajo del nivel de ingreso de pobreza. El gobierno federal administró el financiamiento para la MDTA a través de 12 oficinas regionales del Departamento del Empleo de los Estados Unidos. Las localidades solicitaban financiamiento federal para proporcionar programas de capacitación, lo cual eventualmente, condujo a problemas de duplicación de esfuerzos y a la necesidad de coordinación a niveles más altos del gobierno.

Cuerpo Laboral

El Cuerpo Laboral (Job Corps), establecido en 1964 por el Decreto de Oportunidad Económica (Economic Opportunity Act) es un programa de residencias por un año para jóvenes con desventajas. El cuerpo laboral proporciona instrucción académica de recuperación, capacitación para el trabajo y otros servicios de apoyo. Ha permanecido sin mayores cambios a través de los años.

Decreto de Empleo y Capacitación Integral

Los setentas trajeron enfoques más integrales para resolver los problemas de aquellos con desventajas económicas. Una corriente hacia la descentralización de los programas de empleo transfirió la autoridad en la toma de decisiones del gobierno federal hacia las autoridades estatales y locales. La autoridad tal y como se define en la legislación y en la regulación con frecuencia incluye la responsabilidad para las actividades de diseño, implementación y evaluación de las actividades del programa.

El Decreto de Empleo y Capacitación Integral (Comprehensive Employment and Training Act, CETA) de 1973 introdujo el concepto de un comité consultivo local para asegurar que el interés público local guíe la planeación del programa. La membresía y papel del consejo industrial privado (PIC, por sus siglas en inglés) fueron establecidos en las regulaciones y en algunas localidades la representación fue “garantizada” mediante circunscripciones de educación y trabajo. La capacitación para el trabajo de la CETA fue dirigido a personas con desventajas económicas, receptores de asistencia social y jóvenes con desventajas.

Decreto de alianza para la capacitación laboral

Bajo la administración del presidente Reagan en los ochentas, la capacitación para el trabajo financiada públicamente se reorientó a atender las necesidades de los empleadores del sector privado. La capacitación de habilidades en aulas fue identificada como una de las principales debilidades de los programas anteriores, debido a que continuamente no era el tipo de capacitación deseado por los empleadores locales.

El Decreto de Asociación para la Capacitación Laboral (Job Training Partnership Act, JTPA) de 1982 limitó las opciones de capacitación a las habilidades que eran demandadas por los empleadores locales. LA JTPA también incrementó la proporción de miembros del sector privado en los comités consultivos locales de capacitación para el trabajo como medida para asegurar que sus intereses fueran tomados en consideración. La evaluación fue una parte integral del programa, la cual se decía estaba orientada a dirigir el rendimiento a través de un sistema de indicadores del desempeño por participante para la tasa de reemplazo y las

ganancias. En respuesta a los despidos masivos asociados con la reestructuración económica de las empresas americanas durante los ochentas, la capacitación bajo la JTPA se orientó a los trabajadores desplazados, además de aquellos con desventajas económicas y los receptores de asistencia social.

Decreto de Inversión en Mano de Obra

Para finales de los noventas las condiciones económicas habían mejorado hasta el punto de pleno empleo en la mayor parte de los Estados Unidos. Los más de 30 años de búsqueda para encontrar formas de reducir la pobreza a través de políticas para el empleo evolucionaron hacia un nuevo enfoque que transfiere la responsabilidad del gobierno hacia los individuos, y desvió la autoridad del gobierno federal hacia los estados. Reemplaza el énfasis en la capacitación en habilidades para sacar a una familia de la pobreza, por un énfasis en la colocación laboral que redujera rápidamente el costo de los pagos de apoyos públicos.

Para reflejar estos cambios en la política hacía la autosuficiencia y el control local, el Congreso aprobó el Decreto de Inversión en Mano de Obra (Workforce Investment Act, WIA) en 1998 para reemplazar a la JTPA. La WIA reformó los programas federales de capacitación para el trabajo y creó un amplio sistema de inversión en la mano de obra. El sistema reformado pretende estar enfocado al cliente; para ayudar a los individuos a tener acceso a las herramientas que requieren para gestionar sus carreras a través de información y servicios de alta calidad, y para ayudar a los empleadores a encontrar trabajadores calificados. El nuevo énfasis de la WIA está en “trabajar primero”. En otras palabras, el trabajo es la mejor capacitación. Si los trabajos no están disponibles, la capacitación será principalmente personalizada para servir a las necesidades del empleador, la capacitación en el lugar de trabajo y la capacitación de corto plazo en habilidades fundamentales.

Las principales innovaciones traídas por la WIA son: 1) centros profesionales de trabajo donde todos los programas de empleo y entrenamiento están integrados en una sola ubicación física, 2) cuentas individuales de capacitación que actúan como vales para quienes buscan trabajo y requieren mejorar sus habilidades para tener éxito en el mercado laboral, 3) acceso universal a servicios esenciales para el empleo con acceso sucesivo y más restringido a servicios intensivos y capacitación, y 4) rendición de cuentas monitoreada a través de indicadores de desempeño. Una característica importante de WIA para las áreas locales es el incremento en el control del sector privado sobre el uso de los fondos para la capacitación. Los Comités para el Desarrollo de la Mano de Obra (Workforce Development Boards, WDBs) tendrán una mayoría significativa de miembros de la comunidad de empleadores. La atención hacia aquellos con mayor dificultad para reemplearse y el monitoreo para seguimiento de los resultados fueron conservados de la JTPA.

¿Quién recibe la capacitación para el trabajo?

De acuerdo a las estadísticas comparativas de la OCDE, los Estados Unidos gasta cerca del 0.04 por ciento del PIB en programas de capacitación para el trabajo. Como se muestra en la Tabla 11.3, esta proporción es baja en comparación con otras naciones industrializadas, ubicando a los Estados Unidos entre los cinco países (de 25) con menor inversión de los incluidos en la lista. La capacitación para el trabajo constituye el 26.7 por ciento del gasto total en EU para los programas activos del mercado laboral (ALMPs, por sus siglas en inglés). Este porcentaje es comparable al de Alemania y superior al de Francia y al del Reino Unido.¹

Tabla 11.3 **Gasto gubernamental en capacitación para el trabajo como porcentaje del PIB en 25 países de la OCDE, 2000**

	Como porcentaje del PIB			Capacitación como porcentaje del gasto	
	Entrenamiento	ALMPs	LMPs	ALMPs	LMPs
Dinamarca	0.84	1.55	4.51	54.2	18.6
Finlandia	0.35	1.07	3.29	32.7	10.6
Alemania	0.34	1.23	3.12	27.6	10.9
Suecia	0.31	1.38	2.72	22.5	11.4
Holanda	0.30	1.57	3.65	19.1	8.2
Portugal**	0.30	0.51	1.34	58.8	22.4
España	0.29	0.84	2.18	34.5	13.3
Francia	0.28	1.36	3.12	20.6	9.0
Bélgica*	0.25	1.36	3.70	18.4	6.8
Nueva Zelanda	0.18	0.55	2.17	32.7	8.3
Austria	0.17	0.49	1.58	34.7	10.8
Canadá	0.17	0.51	1.49	33.3	11.4
Grecia**	0.17	0.35	0.83	48.6	20.5
Italia*	0.12	0.63	1.28	19.1	9.4
Corea	0.09	0.46	0.55	19.6	16.4
Suiza	0.09	0.48	1.05	18.8	8.6
Noruega	0.08	0.77	1.16	10.4	6.9
Hungría	0.07	0.40	0.88	17.5	8.0
Reino Unido	0.05	0.36	0.94	13.9	5.3
México*	0.04	0.08	0.08	50.0	50.0
Estados Unidos	0.04	0.15	0.38	26.7	10.5
Japón	0.03	0.28	0.82	10.7	3.7
Australia	0.02	0.45	1.50	4.4	1.3
República Checa	0.02	0.22	0.52	9.1	3.9
Polonia	0.01	0.15	0.96	6.7	1.0

* 1999.

** 1998.

Donde PIB es Producto Interno Bruto, ALMP son políticas activas para el mercado laboral y LMP son políticas para el mercado laboral.

No se encuentran disponibles los datos de los siguientes países de la OCDE: Islandia, Irlanda, Luxemburgo, República Eslovaca y Turquía.

Fuente: OCDE (2001).

La Tabla 11.4 muestra, tanto el gasto público, como el privado en capacitación para el trabajo. La capacitación para el trabajo financiada por el gobierno comprende cerca del 11 por ciento de los 68 mil millones de dólares gastados para dicho propósito en el ejercicio fiscal de 2001. El gobierno federal fue responsable de casi el 90 por ciento del gasto público. De los 6.4 mil millones de dólares de gasto federal en capacitación para el trabajo, el 39.6 por ciento se dirigió a los adultos con desventajas y a los trabajadores desplazados, el 43.3 por ciento a programas para jóvenes (cuerpos laborales y otros), el 6.9 por ciento a servicios comunitarios de empleo para trabajadores mayores, y el 2.1 por ciento a trabajadores que hayan sido afectados por los cambios en las pautas del comercio internacional.

Tabla 11.4 **Estimación del gasto en programas públicos de capacitación para el trabajo en Estados Unidos en el año fiscal 2001**

Miles de US dólares

Programas	Financiamiento federal \$ US	Participación del financiamiento federal (%)	Financiamiento estatal complementario	Financiamiento estatal con requisitos de particulares para 1998	Financiamiento por el empleador	Total de financiamientos
Acciones para los trabajadores adultos y desplazados	\$2 540 040	39.6				
Acciones para los jóvenes	\$1 377 965	21.5				
Cuerpos laborales (jóvenes)	\$1 399 148	21.8				
Programas Nacionales	\$528 150	8.2				
Otros programas (no-WIA)	\$4 500	0.1				
Capacitación TAA	\$94 400	1.5				
Capacitación NAFTA	\$37 150	0.6				
Servicio comunitario de empleo para americanos mayores	\$440 200	6.9				
Financiamiento total	\$6 421 553	100.0	\$276 621	\$593 191	\$60 700 000	\$67 991 365
Porcentaje del financiamiento total	9.4%		0.4%	0.9%	89.3%	100.0%

WIA — Decreto de Inversión en mano de obra (Workforce Investment Act)

TAA — Asistencia para la adaptación comercial (Trade Adjustment Assistance).

NAFTA-Tratado de Libre Comercio de América del Norte (North American Free Trade Act).

Fuente: Wandner, Balducci y Spickard (2001).

Los antecedentes característicos para los participantes en los tres principales programas de empleo y capacitación financiados federalmente se resumen en la Tabla 11.5. Las diferencias en las características de los participantes dentro de los diversos programas son consistentes con los grupos a los cuales se tiene la intención de atender. Ambos programas JTPA Título II-A y Título-III estaban dirigidos a adultos, pero el Título II-A era para personas con desventajas definidas

con ingresos por debajo de la línea de pobreza, mientras que el Título III era para trabajadores desplazados a quienes no les era posible encontrar trabajo y requerían capacitación adicional. En consecuencia, como se muestra en la tabla, un mayor porcentaje de participantes en programas de capacitación del Título II-A eran receptores de asistencia social y era menor el porcentaje de graduados de la preparatoria que los del programa de capacitación del Título III. Más mujeres participaron en el Título II-A que en el Título III. En los programas de capacitación para los jóvenes (JTPA Título II-C) obviamente tienen preponderancia los participantes que no han concluido la preparatoria.

El final de la Tabla 11.5 proporciona información bruta de los resultados de los participantes en los tres principales programas financiados por la JTPA. La incorporación laboral fue del 68 y 69 por ciento para los programas para trabajadores adultos y desplazados, respectivamente; mientras que fue del 47 por ciento para los programas para jóvenes. Para los jóvenes se lograron también proporciones considerables de mejora del empleo o de la competencia; lo cual, la JTPA también considera como un éxito. Entre aquéllos que se incorporaron a un trabajo al terminar el programa, las tasas de ingresos fueron estimadas en 8.75, 7.07 y 11.95 dólares por hora para adultos, jóvenes y trabajadores desplazados, respectivamente.

Tabla 11.5 Características y resultados de los participantes de la capacitación de la JTPA, año fiscal 1999

Características	Adultos Título II-A	Jóvenes Título II-C	Trabajadores desplazados Título III
Número de participantes en el programa	133 744	58 548	189 794
Género: mujeres (%)	65	58	54
Edad: 14 a 15 (%)		7	
Edad: 16 a 21 (%)		93	
Edad: 22 a 54 (%)	97		89
Edad: más de 55 (%)	3		11
Educación: menor a la preparatoria (%)	22	71	11
Educación: preparatoria (%)	56	26	50
Educación: superior a la preparatoria (%)	22	3	39
Raza: negra (%)	35	34	19
Raza: de origen hispano (%)	16	23	13
Raza: caucásico (%)	43	38	62
Individuos discapacitados (%)	7	12	2
Receptores de asistencia social (%)	26	19	2
Exdelincuentes (%)	18	13	5
Receptores de seguro de desempleo	10	1	69
Personas que han agotado su seguro de desempleo (%)	3	1	5
Veteranos (%)	6		11
Resultados:			
Tasa de incorporación laboral (%)	68	47	69
Salario promedio: dólares por hora (\$)	8.75	7.07	11.95

Evaluación de la capacitación para el trabajo en los Estados Unidos

Las dos técnicas más populares para la evaluación de los programas de capacitación para el trabajo son *el monitoreo del desempeño*, el cual escudriña los resultados brutos, y *la estimación del impacto neto*, que evalúa el incremento en el valor por una intervención. Cada uno de estos enfoques tiene ventajas y desventajas [Barnow y Smith (2002), King (2002)]. La orientación de este artículo es sobre la estimación del impacto neto del alcance de la capacitación para el trabajo. Las estimaciones del incremento en el valor derivado de la aplicación de un programa son la base para los cálculos del beneficio neto dirigidos a estimar el retorno sobre la inversión del gasto gubernamental.

Los impactos netos comparan los resultados promedio de una muestra representativa de participantes del programa con una muestra, elegida apropiadamente, de personas similares que no recibieron los servicios. Se debe tener mucha precaución al conformar el segundo grupo, que también es denominado situación contrafactual.² La diferencia entre ambos grupos en las medidas de los resultados de interés es la estimación del efecto del programa. Para los programas de la JTPA, el Congreso indicó que el éxito “será medido por el incremento en los empleos y las ganancias de los participantes y por la reducción de la dependencia de la asistencia social” [Barnow (1989), p. 117].

Métodos para estimar el impacto de la capacitación para el trabajo

Enfoque para la evaluación

La estimación del impacto neto puede basarse en muestras reunidas a través de diseños experimentales clásicos involucrando el muestreo aleatorio, o a través de estudios cuasi-experimentales que aplican procesos estadísticos para simular el ideal de un experimento. En un experimento con diseño clásico, los participantes y los grupos de comparación son creados mediante ensayos aleatorios. Esto significa que las condiciones experimentales fijadas son repetidas un número suficiente de veces para generar los tamaños de muestra requeridos, y en cada repetición se realiza una muestra aleatoria tanto de los participantes del programa como del grupo de comparación.³

Los cuasi-experimentos intentan imitar a un experimento clásico mediante la creación de los grupos de participantes y de comparación apropiados usando procesos estadísticos en lugar del muestreo aleatorio. Un enfoque conveniente es usar la experiencia previa de los participantes en el programa de capacitación para el trabajo como un grupo de comparación para ellos mismos. El estimador del impacto neto asociado es denominado estimador pre-versus post-participación en el programa. Este enfoque es económico puesto que se

apoya en datos administrativos o en fragmentos de entrevistas promedio realizadas. Esto implica que utilizar a los participantes como su propio grupo de comparación automáticamente ajusta el impacto estimado tanto por diferencias en las características observables como en las no observables. Esta estrategia es aceptable si el presupuesto para la evaluación no puede cubrir otro enfoque, pero existen problemas intrínsecos.

Ashenfelter (1978), al efectuar evaluaciones de programas de reentrenamiento bajo la MDTA en Estados Unidos, encontró evidencia de un descenso en los ingresos antes de la separación del empleo en los trabajadores desplazados. El evaluar el impacto neto del programa sobre las ganancias de la recolocación utilizando el diseño pre-post conducirá a una sobre estimación del impacto neto del programa debido a la presencia del efecto *del descenso de Ashenfelter*: “Ashenfelter dip”.

Un análisis pre-post alternativo es seleccionar un grupo de emparejamiento simultáneo adaptando sobre las características observables. Este enfoque selecciona a un grupo de comparación de personas claramente similares que se hayan quedado desempleadas aproximadamente al mismo tiempo que los participantes que contaron con capacitación para el trabajo pero que no hayan participado en el programa. Para cada participante del emparejamiento para el trabajo, se selecciona a un “gemelo” a través de la correspondencia entre variables como edad, género, raza, logros académicos, ocupación anterior, industria anterior, nivel de ingresos-promedio anterior, última recepción de compensación por desempleo, uso previo de programas de impulso a la mano de obra, y la localización geográfica del lugar de residencia.⁴

Si algunas de las características que diferencian a los participantes de los no participantes no son observables, puede ser posible tomarlas en cuenta indirectamente a través de la propensión de emparejamiento de resultados (propensity score matching). Un modelo de predicción, participe o no alguien en la capacitación para el trabajo, es estimado sobre las características observables, algunas de las cuales explican la participación pero no explican el éxito en la recolocación. Dichos factores son difíciles de conseguir por lo que dichos modelos de participación frecuentemente se identifican utilizando transformaciones no lineales de las variables observables. Una vez que el modelo de participación es estimado sobre el total de la muestra puede ser usado para crear un grupo de comparación mediante la selección de no participantes que haya tenido un resultado de su participación cercano a cada participante.

Técnicas de estimación

Si el muestreo aleatorio ha sido implementado adecuadamente en el contexto de un experimento de campo, entonces el impacto neto puede ser estimado

como una diferencia simple de las medias de los grupos de tratamiento y los de comparación. Después de validar la homogeneidad de los grupos de tratamiento y los de control, las muestras asignadas aleatoriamente pueden ser usadas para calcular las estimaciones del impacto neto por una simple diferencia entre las medias:

$$[1] E(y_p) - E(y_c),$$

donde E es el valor esperado que produce las medias de las variables aleatorias binarias, 'y' es un resultado de interés, y el índice p denota la muestra de participantes en la capacitación para el trabajo y c denota la muestra de comparación. Se utilizan valores estadísticos T para probar la significancia estadística.

En términos, claridad en la conducción de las políticas, una estimación desajustada simple del impacto neto, basada en ensayos aleatorios, es usualmente la de mayor influencia debido a que es fácil de entender.⁵ Un enfoque equivalente es usar el análisis de regresión para comparar los resultados entre los grupos de tratamiento y de comparación. Los impactos del programa pueden ser estimados corriendo el modelo de mínimos cuadrados ordinarios:

$$[2] y_i = a_0 + a_1 P_i + u_i,$$

en una muestra conjunta de miembros del grupo de comparación y de participantes de la capacitación para el trabajo, donde y es el resultado de interés, a_1 es el impacto del programa en el resultado para los participantes de la capacitación para el trabajo, a_0 es el valor medio del resultado para los miembros del grupo de comparación, P es una variable dicotómica con valor de 1 para los participantes de la capacitación para el trabajo y 0 en cualquier otro caso, u_i es un término de error distribuido normalmente y con media cero, e i es el índice que denota a los individuos ya sea de la muestra del grupo de participantes o del grupo de comparación. Las pruebas de significancia de los impactos del programa son simplemente pruebas-t sobre el parámetro a_1 .

Para la mayoría de los enfoques cuasi-experimentales la *regresión ajustada a valores observables* es la técnica de estimación del impacto neto conveniente y satisfactoria. Ésta puede ser aplicada a los grupos de comparación creados utilizando muestras de pre- versus post-participación en el programa, o reuniendo no participantes simultáneos. Éste también es un enfoque útil para utilizar datos generados por un experimento de campo cuando no se obtiene la homogeneidad completa de los grupos. Computacionalmente, el método involucra una extensión simple de la ecuación [2]. En dichos casos, la estimación del modelo:

$$[3] y_i = a_0 + a_1 P_i + b_1 X_{1i} + b_2 X_{2i} + \dots + b_n X_{ni} + u_i,$$

por mínimos cuadrados ordinarios en la muestra conjunta de miembros de los grupos de participantes y de los de comparación se generan estimaciones

del impacto neto de programa.⁶ La ecuación [3] es la misma que la ecuación [2], excepto que las características observables (X_i) de los participantes y no participantes son incluidas.⁷

Este método produce impactos netos del programa ajustados por las características observables. Las estimaciones son llamadas “netas” porque los grupos de comparación y de participantes en el programa son ajustados estadísticamente para remover la heterogeneidad entre las muestras. Esto es, el único factor remanente que contribuye a la diferencia en la medida de los resultados es la exposición al tratamiento del programa. La metodología de estimación cancela todos los otros factores observables que afectan el resultado.

Cuando el grupo de comparación ha sido creado usando un algoritmo de emparejamiento de datos, los impactos netos del programa pueden ser estimados simplemente como una diferencia de medias entre los pares emparejados.⁸ Esto es aplicando la técnica de la ecuación [1] a los datos. Algunos analistas también han aplicado el método de la ecuación [3] a dichos datos, donde la co-varianza (X) incluye tanto las características del emparejamiento como algunas variables adicionales.

El popular enfoque de la diferencia en las diferencias implica preparar datos pre-post tanto de los participantes de la capacitación para el trabajo como del grupo simultáneo de no participantes en el programa.⁹ Para todos los temas en cada uno de los grupos se calcula la diferencia entre el valor del resultado más reciente y el valor de un resultado anterior a la participación en el programa. Esta variable es la primera diferencia. El siguiente paso es aplicar la técnica de la ecuación [1] a los datos en esta variable. De ese modo la segunda diferencia es calculada como el contraste de las primeras diferencias entre los participantes y los no participantes de la capacitación para el trabajo. La motivación para este enfoque es el reconocimiento de que el valor anterior del resultado es una variable exógena que involucra todas las maneras observables y no observables en que los dos grupos difieren. En principio el enfoque de la diferencia en las diferencias automáticamente se ajusta por el efecto *del descenso de Ashenfelter* en las ganancias.

Cuando la selección en los programas no es aleatoria y la participación en un programa se debe tanto a características observables como a las no observables, el impacto del programa no puede ser estimado adecuadamente mediante un modelo de regresión del tipo especificado en la ecuación [3]. Heckman (1979) mostró que la selección de muestras sesgaría los parámetros estimados calculados mediante MCO en una ecuación como la [3] como si una variable importante hubiera sido omitida de la especificación de la ecuación estimada. Él recomendó un modo de crear esta variable omitida al ajustar la ecuación de regresión con

base en factores observables y no observables.¹⁰ La razón para estimar impactos utilizando el procedimiento de selección de muestras de Heckman es la preocupación de que existe algún factor no observable sobre los no participantes en el programa, que tiene características observables similares a los participantes en la capacitación para el trabajo, que sería causa de que los primeros tengan diferente éxito en el mercado laboral en comparación con los participantes de la capacitación para el trabajo, incluso aunque hubieran participado en el mismo programa.

Crterios de los buenos métodos de estimación

La Tabla 11.6 enlista siete criterios para valorar si un método de estimación es bueno o no para el propósito que se tiene. Los siete criterios fueron enlistados en la tabla por un aproximado orden de importancia.

Tabla 11.6 **Elementos de diseño en una evaluación de impacto neto**

<p>Enfoque de la evaluación:</p> <p>Diseño experimental clásico</p> <ul style="list-style-type: none"> ● Ensayos aleatorios <p>Cuasi-experimental</p> <ul style="list-style-type: none"> ● Pre-versus post-participación en el programa ● Emparejamiento por características observables ● Emparejamiento por características observables y no observables <p>Técnicas de estimación:</p> <p>Diferencia de medias</p> <p>Regresión ajustada con base en las variables observables</p> <p>Diferencia en medias sobre muestras emparejadas en pares</p> <p>Diferencia en las diferencias</p> <p>Regresión ajustada basada en variables observables y no observables</p> <p>Criterios para buenos métodos de estimación:</p> <p>Transparencia del método de estimación del impacto</p> <p>Insegado</p> <p>Eficiencia</p> <p>Robustez</p> <p>Insensibilidad a tamaños de muestra pequeños</p> <p>Insensibilidad a pequeñas variaciones del estimador</p> <p>Reproducibilidad</p>

El primer criterio es la *transparencia* de la metodología para estimar el impacto. Por esto entendemos que los usuarios importantes de la evaluación de resultados pueden entender con claridad el método de estimación. Las estimaciones de impacto del programa seguramente serán usadas por funcionarios de alto nivel diseñadores de políticas públicas, quienes estarán relativamente capacitados sobre los métodos de investigación en las ciencias sociales. Otros usuarios estarán menos cómodos con complejos procedimientos estadísticos. Las

estimaciones deben ser de interés de aquellos cuerpos directivos que deciden sobre la asignación del presupuesto nacional para el empleo. Idealmente las estimaciones serán utilizadas también por los administradores de programas en el ámbito regional, agencias de empleo en el ámbito local, e incluso participantes potenciales de la capacitación para el trabajo. Los estimados tendrán mayor peso si es clara la manera en que éstos fueron calculados.

El diseño clásico de experimentación de campo satisface el criterio de transparencia mejor que cualquier otro método. Una simple comparación de las medias entre los grupos de tratamiento y de control puede ser muy persuasiva si se demuestra que ambos grupos fueron seleccionados aleatoria y homogéneamente. Sin embargo, se debe reconocer que las evaluaciones experimentales pueden ser prohibitivamente caras. Un enfoque de emparejamiento por pares para grupos de comparación puede ser casi tan convincente como el experimental si se explica claramente. Menos deseable es un método estadístico complejo que puede ser requerido en situaciones donde las muestras obtenidas de los grupos de personas participantes y de los de comparación son muy diferentes.

Dada una muestra apropiada, un estimador insesgado producirá un estimado del impacto que refleje los beneficios a la población de participantes en la capacitación para el trabajo. Si el estimador es *eficiente*, entonces habrá poca variación en los valores del impacto estimado del programa entre muestras alternativas. Si el estimador es *robusto*, no será afectado grandemente si algunas observaciones tienen valores lejanos a la normal.¹¹ Tener un estimador que sea relativamente *insensible a tamaños de muestra pequeños* permite un rango más amplio para análisis de subgrupos demográficos y con las características del diseño del programa.

La *insensibilidad a pequeñas variaciones del estimador* se refiere a la habilidad del analista para controlar por factores observables y no observables al momento de calcular el impacto neto. En la práctica es imposible controlar todos los factores. Con un estimador que sea insensible a pequeñas variaciones, al controlar los factores más importantes, se aislarían otras influencias.

La *reproducibilidad* es una característica deseable en cualquier investigación científica. Las bases de datos deben estar debidamente documentadas y la metodología estadística debe ser bien conocida de manera que otro investigador observando evidencia similar llegará a una conclusión similar sobre los parámetros de interés de la población.

Las dos consideraciones que están notablemente ausentes de nuestra lista de criterios para obtener buenos estimadores son la oportunidad y la efectividad en el costo. Estas dos son consideraciones prácticas para la implementación de un enfoque de evaluación. Los usuarios de la información de la evaluación de los programas tienen plazos de decisión que pueden influir en la elección de un

enfoque, y los mismos tienen presupuestos más o menos limitados para hacer las evaluaciones las cuales también pueden influir en la decisión. La principal diferencia al escoger el enfoque de la evaluación es la elección entre experimental versus cuasi-experimental.

El estimador por diferencia simple de medias que es posible, dado que el enfoque experimental de la evaluación, satisface todos nuestros criterios para un buen estimador. Desafortunadamente, el tiempo y presupuesto no son siempre suficientes para acoger el enfoque experimental. De manera más frecuente los datos para la evaluación son reunidos mediante un diseño cuasi-experimental y las muestras de participantes y de comparación son elegidas ya sea con un diseño pre-post o un enfoque simultáneo. Con el enfoque simultáneo pueden darse amplias correspondencias *ex ante* sobre características observables o bien pueden darse por una simple elegibilidad de escenarios impuestos. En los diseños cuasi-experimentales el último enfoque es el más general. Las muestras simultáneas con afluencia de escenarios de elegibilidad aplicados permitirán aplicar la mayoría de las técnicas de estimación, y por lo tanto seleccionar la mejor aplicando los criterios que se han expuesto.

Aspectos prácticos al evaluar la capacitación para el trabajo

El aspecto más práctico de cualquier diseño de evaluación es el plan para la recolección de los datos. Este componente, que puede ser denominado el diseño de la muestra, tiene cinco elementos esenciales: 1) establecer el tamaño de la muestra, 2) determinar la cobertura geográfica de la evaluación, 3) establecer procedimientos para tomar la muestra, 4) implementación de la encuesta, y 5) examen preliminar de los datos.

Tamaño de la muestra

Las muestras de los participantes y no participantes deben ser del tamaño suficiente para asegurar la precisión las estimaciones deseadas de los impactos. Las muestras grandes permiten detectar los efectos más pequeños de los programas mientras que simultáneamente incrementa la precisión de las estimaciones medidas por los errores de muestreo.¹² Los tamaños de muestra para las evaluaciones de impacto neto deberán ser lo suficientemente grandes para medir confiablemente efectos del tamaño que sean de interés para los diseñadores de políticas. Por ejemplo, si el programa fuera inaceptable debido a que genera menos de 15 puntos porcentuales de incremento en la probabilidad de recolocación, entonces el tamaño de la muestra debe ser establecido para detectar un impacto al menos tan pequeño. El factor restrictivo al recolectar muestras suficientemente grandes para asegurar estimaciones precisas para cualquier efecto esperado es el costo de recolectar los datos.

Tabla 11.7 Aspectos prácticos para la evaluación de la capacitación para el trabajo

Preliminares:

Diseñar la muestra
 Seleccionar aleatoriamente muestras de personas para los grupos de participantes y de comparación
 Extraer registros de archivos administrativos existentes sobre las muestras seleccionadas
 Preparar un archivo de datos para un análisis preliminar de las muestras seleccionadas
 Preparar listas de nombres para las entrevistas

Levantamiento de la encuesta:

Elaborar encuestas para prueba piloto
 Revisar las encuestas y definir los formatos finales para registrar las respuestas de las encuestas
 Preparar las encuestas en el formato requerido para las entrevistas
 Preparar un manual de entrenamiento para los encuestadores sobre la aplicación de la encuesta
 Designar a los encargados de las encuestas en las principales regiones geográficas
 Organizar equipos de encuestadores para realizar las entrevistas
 Guiar la capacitación de los encuestadores
 Guiar las entrevistas con protocolos establecidos
 Entregar los cuestionarios completados para capturar los datos

Procesamiento final de los Datos:

Verificar errores, corregir y capturar los datos codificados en archivos informáticos
 Preparar los archivos informáticos para el analizar los datos
 Entregar los archivos a los analistas de datos
 Corregir los archivos de datos con base en elementos del análisis de datos

Estimación de Impacto

Fuente: Retomada de la tabla 4.13 en O'Leary, Nosporova y Samorodov (2001, p.106).

Selección del sitio

Todas las evaluaciones a programas de capacitación son realizadas en alguna localidad, dado que la capacitación es de por sí para una ubicación específica. Al dirigir una evaluación de impacto neto con entrevistas de seguimiento, deben estar dadas las consideraciones prácticas de un tipo de encuesta. Las encuestas telefónicas involucran menores consideraciones geográficas que las encuestas casa por casa. Para las encuestas en persona, los costos de transporte son el principal componente de los costos de la encuesta, particularmente en áreas con población rural significativa. Un requerimiento estándar en las evaluaciones en los Estados Unidos es que los sitios deben seleccionarse de modo que sean representativos del estado o región donde se lleva a cabo la evaluación.

Selección de la muestra

En experimentos de campo diseñados clásicamente, la aleatoriedad es aplicada al momento de asignar el programa dentro de una oficina de empleo local. Idealmente, la selección de la muestra también opera como un proceso de flujo para las evaluaciones cuasi-experimentales. Tanto los miembros que participan en la capacitación para el trabajo como los del grupo de comparación deben estar en la misma situación del mercado laboral y ambos deben tener la misma elegibilidad para participar en el programa.

Para las evaluaciones cuasi-experimentales, las extracciones de las muestras de los participantes en el programa y del grupo de comparación de los sistemas de datos de las oficinas gubernamentales de empleo permiten economizar la extensión de cualquier encuesta de seguimiento necesaria. En la mayoría de los países será posible obtener información valiosa de los sistemas de datos computarizados de las oficinas de empleo utilizados para administrar el programa. Dichos sistemas, con frecuencia, pueden proporcionar información demográfica básica confiable sobre aspectos como edad, género, logros educativos y composición familiar. Otro tipo de información como es la ocupación habitual y la industria de empleo, promedio de ingresos mensuales anteriores, y la incorporación previa a programas gubernamentales de empleo o de apoyo al ingreso, está potencialmente disponible. Recopilar dicha información de los registros administrativos mejora la exactitud y acorta las encuestas de seguimiento para enfocarse en las experiencias de recolocación.

Implementación de la encuesta

Los investigadores de encuestas han encontrado que se requiere de una tasa de respuesta de alrededor del 80 por ciento para asegurar que los datos reunidos reflejan a la población del esquema muestral. Un enfoque rápido para llevar a cabo el trabajo de campo de la encuesta es contratar a una compañía privada de encuestas. Utilizar a una compañía privada independiente que sea un observador desinteresado agrega también un mayor nivel de objetividad al proceso, lo que puede ayudar a validar los hallazgos. Sin embargo, el costo por encuesta completada usando dicho proveedor externo puede ser prohibitivo para algunos programas, particularmente si son financiados por gobiernos locales. Un enfoque alternativo es desarrollar la capacidad de investigación de la encuesta al interior del sistema de los administradores locales de la capacitación para el trabajo. Algunas entidades administrativas locales ya tienen la responsabilidad de realizar el seguimiento con participantes anteriores para medir los resultados, y esta capacidad puede ampliarse para las evaluaciones.

El plan debe requerir un sondeo completo de todos los nombres listados en el primer intento de contacto. El número de entrevistas completadas en la primera ronda debe entonces ser comparado con la muestra objetivo. Si el número queda significativamente corto respecto al objetivo, entonces debe hacerse un segundo intento para convocar a todos aquellos sujetos no contactados en la primera vuelta. Este proceso debe efectuarse hasta que el objetivo de la muestra sea logrado. Mediante este enfoque la muestra final estará apenas por debajo o un poco por encima del número del diseño. Esta estrategia mantiene la representatividad de la muestra de modo que cada persona en el marco del muestreo original tiene igual oportunidad de estar en la muestra final.

Un plan de sondeo que no debería seguirse en lo absoluto es detener todas las entrevistas inmediatamente después de haber completado el número de entrevistas fijado. Siempre que se inicia una ronda de entrevistas debería completarse. Interrumpir una ronda disminuye la representatividad de la muestra.

Al comparar el costo de los enfoques alternativos para el trabajo de encuestas de campo deben considerarse todos los aspectos del proyecto. Una vez que la muestra ha sido extraída, se requieren muchos pasos antes de que un archivo final de datos este disponible para el análisis. En su secuencia cronológica usual, los principales pasos son: 1) entrenar a los encuestadores, 2) realizar pruebas piloto con los cuestionarios, 3) revisar los cuestionarios, 4) imprimir los cuestionarios, 5) distribuir las listas de direcciones para entrevistar y las copias de los cuestionarios a los encuestadores, 6) llevar registro de los intentos realizados para completarlas, 7) supervisar la exactitud e integridad, 8) capturar en computadora los datos reunidos en la encuesta, y 9) verificar errores de los datos de la encuesta en el archivo informático.

Examen preliminar de los datos

En un diseño cuasi-experimental, los datos para la evaluación provienen usualmente de dos fuentes: registros administrativos y encuestas de seguimiento. El primer paso en la evaluación es examinar los datos para su integración y corregir los errores. Deben imprimirse y examinarse resúmenes con información de todas las variables. El resumen informativo debe incluir la media, el mínimo, el máximo y el recuento de observaciones con los datos faltantes. Para todos estos debe verificarse que queden comprendidos en el rango de valores aceptables. Posteriormente se aplican procedimientos aceptados para corregir datos extremos irracionales y valores faltantes.¹³

Adicionalmente, es necesario verificar la aleatoriedad en la creación de los grupos de participantes en el programa y de comparación.

Ejemplos de evaluaciones federales de capacitación para el trabajo

Cada uno de los principales programas federales de capacitación para el trabajo ha sido evaluado de una forma u otra. La evaluación bajo la MDTA examinó resultados brutos basados en encuestas a participantes anteriores. Se realizaron mejoras cuando CETA realizó evaluaciones mediante la inclusión de un grupo de comparación. Las evaluaciones de la JTPA dieron un paso adelante al adoptar un diseño con muestreo aleatorio. Los cuerpos laborales estuvieron sujetos tanto a enfoques de evaluación cuasi-experimentales como de muestreo aleatorio. Actualmente está en ejecución una evaluación de la WIA basada en un enfoque cuasi-experimental utilizando datos administrativos. Dado que la JTPA

y los cuerpos laborales son los dos únicos programas que utilizaron diseños de muestreo aleatorio como diseño de evaluación. La mayoría de lo que sabemos sobre la efectividad de los programas de entrenamiento para el trabajo se dedujo de las evaluaciones de estos dos programas. También hemos hecho notar en la descripción de estas evaluaciones, el papel que juegan las evaluaciones en el proceso político de la reautorización legislativa.

Decreto de desarrollo y entrenamiento de la mano de obra

La MDTA fue el primer intento federal para apoyar a los trabajadores desplazados a reincorporarse a un empleo mediante el reentrenamiento en habilidades laborales. Reflejaba la preocupación principal al momento de perder el trabajo debido al cambio tecnológico (Leigh, 1990). Entre septiembre de 1962 y septiembre de 1967, más de 601 000 personas fueron inscritas en programas de reentrenamiento organizados por las localidades que se beneficiaron directamente de oficinas regionales del Departamento del Trabajo de los EU.

La evaluación de la MDTA se apoyó en encuestas de seguimiento de los participantes en el programa. Al momento de la principal evaluación de la MDTA, 74 000 participantes aún estaban involucrados en los programas de reentrenamiento y el 30 por ciento los había abandonado. Entre los participantes preeliminares, el 90 por ciento se había reincorporado durante el siguiente año a la capacitación, y el 77 por ciento estaba empleado al momento de aplicarse la última encuesta de seguimiento (Mágnun 1969, p. 81). Sin embargo, como se mencionó en la sección anterior, estas estimaciones brutas de los resultados no son indicadores confiables del impacto neto de los programas.

Las previsiones al término de la legislación que diseñó el MDTA terminaron con el programa en 1969 y obviamente la evidencia de la evaluación del programa no evitó que se diera por terminado. La principal razón para desaparecer el MDTA fue la estructura administrativa a través de la cual la autoridad de las entidades políticas estatales y locales fue excluida mediante contratos federales otorgados directamente a proveedores locales de servicios, y la entrega de dichos servicios era duplicada a nivel local.

CETA

El Decreto de Empleo y Capacitación Integral (CETA, por sus siglas en inglés) de 1973 fue el primer programa capacitación para el cual el Departamento del Trabajo de los EU desarrolló una base de datos específicamente orientada a la evaluación del programa (Leigh 1990, p. 10). Fue denominada Encuesta Continua Longitudinal de la Mano de Obra (CLMS, por sus siglas en inglés) y contenía datos sobre los participantes en el programa, sobre los miembros del grupo de

comparación extraídos de la encuesta nacional de la fuerza laboral (Encuesta de la Población Actual), y sobre los ingresos de todos los individuos registrados en el seguro social nacional (Seguridad Social). Los estudios de evaluación se facilitaron con la creación de la CLMS, a pesar del hecho que los programas CETA estaban dirigidos a los individuos de bajos ingresos mientras que la encuesta de la fuerza laboral representaba a la nación.

Tres resultados principales surgieron de las 11 primeras evaluaciones a CETA (Leigh 1990, p. 11). Primero, no hubo un impacto medible en el empleo o en el ingreso para hombres. Por otro lado, el impacto para las mujeres, fue positivo y significativo. Segundo, la capacitación en el lugar de trabajo era usualmente más efectiva que la capacitación en el aula. Finalmente, el rango de las estimaciones de impacto fue bastante amplio, a pesar del hecho de que todos los analistas usaron los mismos datos de la CLMS.

El CETA expiró en 1982. Sin embargo, su terminación fue resultado de cuestiones alrededor de la administración del programa y malos manejos de los fondos más que por la evidencia de sus efectos en los trabajadores.

JTPA

La aprobación de la JTPA fue resultado de verdaderos compromisos ideológicos y partidarios entre los demócratas liberales y los republicanos conservadores. Muchas características del proyecto de ley reflejaron los compromisos requeridos para su aprobación y decreto, incluyendo el mandato del Congreso de los EU para una primera evaluación nacional del programa. Para asegurar la objetividad en la evaluación del impacto neto, el Congreso insistió en que la evaluación debía apoyarse en métodos experimentales de campo con una muestra aleatoria de los grupos sujetos a la capacitación y los de comparación en 16 sitios a lo largo del país. Orr *et al.* (1996, p. 109) reportó que la capacitación para los adultos con desventajas económicas resultó en el incremento de salarios del 11 por ciento en mujeres y un 6.7 por ciento en hombres. Para ambos géneros, el incremento en los ingresos se debió principalmente a que aumentaron las horas trabajadas. Tanto las mujeres como los hombres experimentaron beneficios netos positivos, y el beneficio neto para la sociedad de ambos géneros fue apenas superior a 500 dólares por participante (Orr *et al.*, 1996, p. 189).

Cuerpos Laborales

La primera evaluación de los Cuerpos Laborales fue cuasi-experimental (Mallar *et al.*, 1980) y halló efectos positivos modestos sobre el empleo y los ingresos semanales, pero ningún impacto en las tasas salariales por hora. Un estudio reciente fue llevado a cabo bajo un diseño clásico de experimentación de

campo (Burghardt et al., 2001); este nuevo estudio encontró que la participación en los Cuerpos Laborales tuvo como resultado incrementos significativos en los ingresos de los jóvenes con desventajas. “Incluso, los incrementos en los ingresos, el progreso en la educación y otros cambios positivos fueron encontrados en la mayoría de los grupos de participantes y se espera que persistan a medida que envejeczan”.¹⁴ Se reportó que el estudio más reciente fue un instrumento para salvar a los Cuerpos Laborales de ser eliminados de las recientes rondas presupuestales.

WIA

Aunque la WIA fue decretada en 1998, la implementación completa de la WIA no inició hasta julio del año 2000. En consecuencia, sólo recientemente ha pasado suficiente tiempo para que el impacto total del programa pueda verse. Apenas se han iniciado los trabajos en un estudio cuasi-experimental del impacto neto de la capacitación en siete estados (Hollenbeck 2002). Los estados involucrados son Washington, Missouri, Illinois, Maryland, Florida, Georgia y Texas. También está en progreso un experimento de campo para evaluar los vales de capacitación bajo la WIA (Decaer y Perez-Johnson, 2002).

Como precursor de esta evaluación más profunda, Hollenbeck utilizó el mismo método para evaluar los programas de capacitación en el estado de Washington. Hollenbeck (2002) utilizó el enfoque no-experimental de agrupamiento estadístico para evaluar programas de desarrollo de la mano de obra en el estado de Washington.¹⁵ Empleando datos administrativos reunidos por diversas agencias estatales, Hollenbeck tomó a los solicitantes de los servicios estatales de empleo como la base de individuos para emparejarlos contra los participantes en los programas de capacitación. En seguida, aplicó un agrupamiento estadístico a nivel de individuo por individuo para hallar a las personas cuyas características observables se emparejaran de manera más cercana con las de los participantes del programa. Los impactos netos fueron determinados entonces al comparar los resultados de los individuos que participaron en los programas de capacitación con las contrapartes de los datos del servicio de empleo que encajaron, quienes nunca participaron en programa alguno. Mediante el uso de este método, Hollenbeck evaluó nueve programas de capacitación para el trabajo en el estado, incluyendo aquéllos asociados con las universidades comunitarias. Un beneficio de este enfoque es que los datos ya están reunidos por las diversas agencias y está disponible fácilmente para cualquier evaluación, eliminando la necesidad de recolectar datos adicionales que es típicamente el caso para los experimentos de muestreo aleatorio.

¿Qué funciona para quién?

Las evaluaciones de la capacitación para el trabajo dentro de los Estados Unidos han sido efectuadas por tipo de servicio y categoría del participante objetivo de los programas financiados federalmente. Los principales grupos objetivo incluyen a adultos y jóvenes con desventajas y a trabajadores desplazados. Dentro de estos grupos objetivo han surgido diferentes resultados por género y dependiendo si la capacitación para el trabajo se efectúa en el aula o de forma práctica.¹⁶

Los resultados de dos evaluaciones nacionales llevadas a cabo como experimentos de campo involucrando muestreo aleatorio son representativos del esquema común de las estimaciones de impacto de la capacitación para el trabajo. Nosotros ilustramos este tipo de hallazgos con la evidencia de dichas evaluaciones: la evaluación nacional de la JTPA (Orr *et al.*, 1996) y la evaluación de los Cuerpos Laborales (Burghardt *et al.*, 2001). Vale la pena resaltar que la evidencia de los experimentos de la JTPA y la de los Cuerpos Laborales es consistente con otros resultados, ya que ambas evaluaciones aplicaron y definieron bien un grupo en situación opuesta. Cuando el tratamiento es la capacitación para el trabajo, la mayoría de las evaluaciones requieren que el grupo de control sea excluido de todos los servicios. Las evaluaciones de la JTPA y los Cuerpos Laborales definen un grupo contrafactual que permite que el grupo de control pueda haber recibido cualquier cantidad de servicios diferentes que la capacitación para el trabajo financiada públicamente que está siendo evaluada. Este diseño proporciona una mejor validez externa, lo que significa que la evidencia del experimento se aplica de una manera más natural al contexto del mundo real.

Entre los adultos con desventajas, el impacto medible para los hombres tiende a ser pequeño. Por otro lado, el impacto para las mujeres tiende a ser positivo y significativamente mayor que en el caso de los hombres. Entre las mujeres receptoras de asistencia social el impacto en los ingresos también es significativo y positivo. Para los jóvenes, la capacitación para el trabajo de corto plazo tiende a no ser efectiva; mientras que a mayor duración, los Cuerpos Laborales de capacitación parecen ser más efectivos para los jóvenes con desventajas. Más aún, la capacitación en el sitio del trabajo (OJT, por sus siglas en inglés) es usualmente más efectiva que la capacitación en el aula (CT, por sus siglas en inglés).

La evidencia de la evaluación nacional de la JTPA se resume en la Tabla 11.8. Ésta muestra los impactos en los ingresos en un periodo de 30 meses posteriores al ingreso en el programa. Los hombres adultos que participaron en la capacitación del programa JTPA obtuvieron un promedio de 1 599.00 dólares en ingresos y tendieron a beneficiarse más de la capacitación para el trabajo en el lugar de trabajo (OJT) que en el aula (CT).

Tabla 11.8 **Impacto de la capacitación para el trabajo por programa y grupo demográfico**

Programa por grupo demográfico	Impacto medio en los ingresos (usd)		Error estándar
Hombres adultos			
JTPA	\$1 599	*	86 565
OJT	\$2 109		133 535
CT	\$1 287		158 282
Mujeres adultas			
JTPA	\$1 837	**	52 525
OJT	\$2 292	**	102 323
CT	\$630		67 070
Mujeres adultas con asistencia			
JTPA OJT	\$4 833	**	172 929
JTPA CT	107 777		106 262
Hombres jóvenes			
JTPA OJT	-\$3 012		222 222
JTPA CT	\$251		191 616
Mujeres jóvenes			
JTPA OJT	-\$579		188 383
JTPA CT	\$839		79 191

* Estadísticamente significativa a un nivel de 90 por ciento en una prueba de dos colas.

** Estadísticamente significativa a un nivel de 95 por ciento en una prueba de dos colas.

Fuente: Orr et al. (1996).

Notas: el impacto en los ingresos es estimado en los 30 meses posteriores al inicio del programa, JTPA es el Decreto de Asociación en el Entrenamiento para el Trabajo (por sus siglas en inglés), OJT es capacitación en el lugar de trabajo bajo el JTPA (por sus siglas en inglés), CT es la capacitación en el aula bajo la JTPA (por sus siglas en inglés).

Los impactos de la JTPA para las mujeres adultas tienden a ser positivos y estadísticamente significativos. El impacto puntual es de 1 837 dólares a los 30 meses, con la OJT significativamente más efectiva que la CT. Para las mujeres receptoras de asistencia social bajo la JTPA, el OJT tuvo un impacto positivo de 4 833 dólares sobre los ingresos, y esto fue significativamente mayor que el impacto del CT en este grupo.

Para los jóvenes, la capacitación de corto plazo ofrecida por la JTPA no fue eficaz; sin embargo, la costosa capacitación en los Cuerpos Laborales de mayor duración, tuvo un considerable impacto positivo. Los Cuerpos Laborales tuvieron beneficios neto por participante a la sociedad de casi 17 000 dólares en la vida de los participantes en el programa (Burghardt et al., 2001). Este resultado para los Cuerpos Laborales de jóvenes se debe a la combinación de varios efectos: horas adicionales de trabajo, tasa salarial mayor y menor asistencia social.

Impacto global de los programas federales de capacitación para el trabajo

Como se mencionó en la sección anterior, el gasto en programas de capacitación para el trabajo proporcionados públicamente representan un pequeño porcentaje del monto total estimado del gasto de los empleadores del sector privado en capacitación a su fuerza de trabajo. Sólo el 10 por ciento del estimado de 68 mil millones de dólares gastados en capacitación durante el año 2001 fueron financiados por el gobierno federal o por los gobiernos estatales. El resto fue financiado por las empresas. Además, estos programas gubernamentales sólo sirvieron a una fracción de los trabajadores con los requisitos necesarios. Por ejemplo, el Reporte de la Oficina de Contabilidad del Gobierno de los Estados Unidos estimó que a finales de los ochenta la JTPA sirvió a no más del 6 por ciento del total estimado de la población que contaban con los requisitos necesarios.¹⁷ Este porcentaje disminuyó aún más cerca de la conclusión del programa a finales de los noventa.

Por lo tanto, se plantea la cuestión hacia el impacto neto del programa en la población objetivo, más allá de su efecto en quienes efectivamente recibieron los servicios. Es una cuestión a escala y del objetivo planteado para el programa. Si el pleno empleo es la meta (según lo medido por el índice de desempleo para una región o una nación), entonces la cuestión es cuándo debería aumentar el programa en términos tanto de cobertura como de intensidad de los servicios, para alcanzar dicho objetivo. La meta de pleno empleo de una región o grupo objetivo puede ser alcanzada, indudablemente con mucho mayor esfuerzo financiero, sin progresar mucho en mover a las personas arriba de la línea de pobreza. Si reducir la pobreza es también un objetivo, entonces es concebible que el programa tenga que expandirse aún más. Por ejemplo, Heckman, Roselius y Smith (1994) estimaron que para reestablecer los niveles salariales de los graduados de la educación media superior de 1989 respecto de los niveles de 1979, se requerirían 212 mil millones de dólares (en dólares de 1989) y otros 214 mil millones de dólares para restaurar el descenso en el nivel salarial de la población masculina que no terminó la educación media superior, asumiendo un 10 por ciento de retorno por la capacitación. Incluso el recuento más generoso del gasto gubernamental anual para la capacitación de los trabajadores palidece en comparación con dicho monto.

Programas de capacitación para el trabajo con financiamiento estatal dirigidos a las empresas

Muchos estados subsidian la capacitación de los trabajadores de planta a través de programas de capacitación personalizados, de acuerdo con las necesidades de los negocios. Estos programas estatales difieren de los programas federales descritos en las secciones previas en que su principal propósito

es promover el desarrollo económico local y regional a través de promover la atracción, creación y conservación de los negocios. La mayoría de los estados financian la capacitación a través de un fondo general, pero algunos dependen de fuentes de financiamiento como los impuestos asociados con el seguro de desempleo o los bonos. En 1999, los 45 estados que ofrecieron dichos programas gastaron cerca de 600 millones de dólares en programas de capacitación personalizados, lo que representa alrededor del 20 por ciento del total de gasto por la WIA en capacitación para adultos y trabajadores desplazados (Moore *et al.*, por publicarse). Desde una perspectiva de desarrollo económico, estos programas estatales de capacitación comprenden el 34 por ciento de los 5.2 mil millones de dólares gastados en el año 2000 por los gobiernos estatales y locales en actividades para el desarrollo económico.

Los programas estatales de capacitación difieren en tamaños y tipos. El gasto en capacitación personalizada y adaptada a las necesidades varía entre 117 millones de dólares en California y 500 millones de dólares en Vermont (Moore *et al.* por publicarse). Mientras que la mayoría de los programas se enfocan en los trabajadores de planta; algunos pocos se dirigen a los trabajadores con desventajas o desplazados. Por ejemplo, el estado de Texas administra ambos tipos de programas. El Fondo para el Desarrollo de Habilidades [Skill Development Fund] apoya a las empresas con la capacitación de su fuerza laboral. En alianza con universidades públicas comunitarias y técnicas y con una agencia de extensión académica, el Fondo para el Desarrollo de Habilidades apoya con financiamiento a programas de capacitación para el trabajo personalizados acordes con las necesidades expresas de las empresas estatales. El Fondo de Autosuficiencia [Self-Sufficiency Fund] se dirige a los receptores de Asistencia Temporal para Familias Necesitadas [Temporary Assistance for Needy Families (TANF), por sus siglas en inglés] o de cupones de comida con el propósito de apoyarlos para obtener capacitación, encontrar un empleo e independizarse de los apoyos financieros del gobierno. Así, mientras que el segundo programa todavía trata de cubrir las necesidades de mano de obra de las empresas, se enfoca más a un grupo de la población objetivo. Cada programa gastó cerca de 12 millones de dólares en el año 2001, y en conjunto sirvieron a más de 450 empresas que ofrecieron capacitación a 15 000 trabajadores. La capacitación que se otorgó a través del primer programa cuesta un poco menos de 1 000 dólares por participante, mientras que la capacitación otorgada a través del segundo programa cuesta más de 3 000 por individuo capacitado.¹⁸

Metodología para la evaluación

Dado que el propósito de los programas estatales de capacitación adaptados a las necesidades es el promover el desarrollo de las economías locales, la evaluación del impacto de estos programas requiere de más elementos que la simple suma de

los efectos individuales de los participantes en un programa específico. También se deben tomar en cuenta las complejas relaciones e interacciones entre los entes económicos locales con el propósito de valorar el impacto de la capacitación en la economía local. En particular, la evaluación debe incluir el efecto de la capacitación en la productividad y en los salarios, así como tomar en cuenta los efectos a escala y de desplazamiento de los programas. Bartik, en su artículo para esta conferencia, examina varios enfoques y concluye que “las evaluaciones rigurosas pueden ser efectuadas a través de experimentos aleatorios, de análisis estadísticos de los usuarios del programa y de grupos de comparación, encuestas y grupos de enfoque, y relacionando modelos econométricos regionales con el impacto fiscal y con modelos del mercado laboral local”.

Sin embargo, es necesario reconocer que las evaluaciones del efecto regional de los programas estatales de capacitación para el trabajo son más difíciles de llevar a cabo y tienden a ser menos precisas que las evaluaciones de programas de capacitación enfocados a trabajadores individuales. Dichas evaluaciones deben considerar indicadores regionales amplios, como son la creación de empleos o la reducción de la pobreza, que resultan de intervenciones relativamente pequeñas. Las evaluaciones rigurosas deben incorporar también efectos a escala y de desplazamiento, los cuales son difíciles de modelar y estimar. Es factible la metodología de muestreo aleatorio, pero es difícil aplicarla a la evaluación de los programas estatales puesto que los programas estatales de capacitación para el trabajo están dirigidos a empresas específicas. Lo más que se puede hacer en muchos casos es comparar los resultados esperados de las empresas participantes en el programa con las no participantes. Pero como se explicó antes, esto añade imprecisiones adicionales a la evaluación. Por lo tanto, quienes tienen prevista una evaluación de programas estatales de capacitación para el trabajo, así como otros programas sociales administrados localmente, enfrentan la posibilidad de evaluar con pocas expectativas acerca de la utilidad del esfuerzo.

En esta sección se considerarán tres tipos básicos de metodologías de evaluación. El primero se apoya en grupos de comparación. El segundo utiliza modelos econométricos para modelar explícitamente las relaciones clave entre la intervención y los resultados esperados. El tercer enfoque, una variación de la segunda metodología, se enfoca en una empresa representativa para esbozar los efectos relativos de los programas sobre la situación financiera de una empresa.

Construcción de grupos de comparación

Dado que los programas estatales de capacitación generalmente se dirigen a empresas, la selección de los grupos de comparación debe sustentarse en firmas y no en individuos. Un ejemplo de este enfoque es el estudio realizado por Holzer et al. (1993), en el que utilizó un enfoque cuasi-experimental para evaluar un

programa en Michigan que proporcionaba por una sola vez subvenciones para las empresas elegibles. Se construyó un grupo de comparación capitalizando el hecho que no todas las empresas que postularon para el programa tuvieron éxito en recibir los recursos. Se compararon los efectos de las empresas que no fueron exitosas al obtener los recursos con aquellas que sí lo fueron (grupo de tratamiento), asumiendo que ambas tenían la motivación para participar en el programa. Sus resultados mostraron que aquellas empresas que recibieron las subvenciones para la capacitación incrementaron significativamente la cantidad de servicios de capacitación que ofrecieron a sus empleados. El estudio encontró también que aquellas empresas que recibieron las subvenciones para la capacitación experimentaron una mejora en la productividad. En un análisis separado se identificó un pequeño efecto del programa de capacitación sobre los salarios y tan sólo un modesto incremento en el empleo de las empresas.

Hollenbeck (2002) utiliza datos administrativos de los programas de empleo en el estado de Washington para efectuar una evaluación cuasi-experimental. El grupo de comparación se forma con participantes emparejados estadísticamente con no participantes de acuerdo con las características observables. Todos los programas de capacitación para el trabajo y los programas de educativos que él evaluó mostraron un impacto neto positivo en el empleo, con excepción de la educación básica para adultos en las universidades comunitarias. Mientras que los programas evaluados por Hollenbeck no incluían capacitación adaptada a requerimientos especiales, este estudio es instructivo para mostrar como los datos administrativos pueden ser utilizados para integrar grupos de comparación.

Factores responsables de influir en el desarrollo económico

Otro enfoque para evaluar el efecto de los programas estatales de capacitación sobre los resultados en el desarrollo económico es examinar los diversos componentes que vinculan a los programas de capacitación con los cambios en la ventaja en el costo de las empresas y así su efecto en la economía local. Estos factores pueden incluir:

1. los efectos del programa sobre los salarios y la productividad individual de los trabajadores;
2. el efecto de los programas en la productividad resultante y en el costo unitario de la mano de obra de las empresas, y
3. subsecuentemente su efecto en el comportamiento local de las empresas y en el crecimiento.

La capacitación puede tener también efectos indirectos. Por ejemplo, las empresas pueden considerar como lugares favorables para ubicarse a las comunidades que invierten en la calidad de la fuerza de trabajo, debido a que

indican que la comunidad valora otros aspectos menos tangibles que contribuyen a crear un clima favorable para las empresas. Más aún, estos efectos deben ser entendidos desde el interior del contexto de las condiciones de los mercados de trabajo locales y con respecto a la escala del programa, esto es, el número de participantes en estos programas con relación al número de empleados en la empresa y el número de participantes potenciales en la comunidad. Se debe considerar también el efecto de desplazamiento de los salarios mayores en los trabajadores y en las empresas, que podría anular cualquier efecto positivo de los programas.

Una vez que se obtienen las estimaciones de los impactos de la capacitación relacionadas con los factores individuales, estas pueden colocarse en un modelo econométrico regional. Dicho modelo es REMI (Regional Economic Models, Inc.), que ha sido presentado en esta conferencia. El modelo permite estimar los efectos directos e indirectos de las políticas de intervención tales como los programas de capacitación.

Una alternativa más sencilla es considerar los efectos del salario y de la productividad en los costos unitarios de mano de obra, y luego determinar si estos programas afectan o no los costos operativos de las empresas lo suficiente para influir en sus decisiones locales.¹⁹

Una ventaja de este enfoque es su simplicidad y transparencia. La desventaja es también su simplicidad, debido a que pueden existir interacciones importantes entre el programa y otros aspectos de la economía local que pasan inadvertidos a menos que sean capturados en un modelo econométrico riguroso.

Una evaluación combinando grupos de comparación y modelos econométricos regionales

Una evaluación del programa de capacitación para el trabajo del estado de California, realizada por Moore, Blake, Phillips y McConaughy (por publicarse), combina el método de grupo de comparación con el de enfoque regional. Los evaluadores integraron a los grupos de comparación, compuestos por empresas similares a aquellas que están recibiendo las subvenciones para la capacitación, y después relacionaron las diferencias en los resultados entre los dos grupos a través de los efectos económicos multiplicadores y de desplazamiento. Los evaluadores aumentaron el análisis del impacto neto con estudios de caso con el propósito de aprender más del porqué en las diferencias de los resultados entre los dos grupos.

El Panel de Capacitación para el Empleo de California (California's Employment Training Panel, ETP) fue diseñado para incrementar la productividad de las empresas existentes en el estado, atraer nuevos negocios y así reducir el desempleo. El programa, en esencia, estableció una alianza entre el estado y las compañías participantes para capacitar a sus trabajadores. El estado reembolsa a los

empleadores el costo directo por capacitar a nuevos reclutas y a los adscritos a la empresa. A cambio, la compañía paga por las instalaciones para la capacitación, cubre los salarios de los trabajadores y el costo de la pérdida en la producción mientras los trabajadores asisten a la capacitación. De 1994 a 1995, California gastó 73 millones de dólares en el ETP con un costo promedio por trabajador capacitado que oscila de 150 a 260 dólares por hora de clase. Los evaluadores encuestaron a las empresas que recibieron el subsidio, y usaron los registros administrativos del Seguro de Desempleo para integrar los grupos de comparación.

Durante el periodo de la evaluación de 1994-95, más de 57 000 trabajadores fueron capacitados. El tipo de contrato ETP para la capacitación más común fue uno en el que el ETP contrató con una sola compañía para proporcionar la capacitación para sus trabajadores. El segundo enfoque más frecuente de contrato ETP fue con agencias de capacitación, como universidades comunitarias, organizaciones privadas de capacitación o asociaciones de comercio para proporcionar tipos específicos de capacitación a trabajadores en empresas elegibles por el ETP. Bajo este arreglo contractual, las agencias de capacitación corren el riesgo de que los reclutados que se capacitan no completen la colocación y retengan el trabajo durante los 90 días requeridos, en cuyo caso el capacitador no recibiría su pago. Los evaluadores encontraron que la empresa típica que recibe la capacitación para sus trabajadores a través del ETP emplean 200 personas y tienen una nómina anual de 7.2 millones de dólares.

Moore *et al.* (por publicarse), estimó los siguientes impactos derivados del programa de capacitación ETP en los empleados de las empresas que recibieron apoyos de la ETP; estos se muestran en la Tabla 11.9. Utilizando un enfoque de la diferencia entre las diferencias, compararon el crecimiento en el número de empleados, salarios pagados e ingresos por empleado de las empresas del ETP con aquellos con que cuentan las empresas del grupo de comparación. Encontraron que las empresas del ETP experimentaron un crecimiento significativo en el número de empleados y salarios pagados que el que tuvieron las empresas del grupo de comparación, pero no encontraron una diferencia significativa en el crecimiento de los ingresos por empleado.

Tabla 11.9 Tasas de Crecimiento de los resultados seleccionados para los trabajadores de las empresas que recibieron capacitación a través del ETP de California con respecto a los de las empresas del grupo de comparación

Resultados	Tasas de crecimiento		Diferencias entre diferencias
Empleados	ETP	14.3%	
	Comparación	-0.8%	15.1%
Salarios pagados	ETP	25.8%	
	Comparación	10.2%	15.6%
Ingresos/Empleado	ETP	11.9%	
	Comparación	11.7%	0.2%

Fuente: Moore, Blake, Phillips y McConaughy (Por publicarse)

Los evaluadores también estimaron que el impacto del ETP en la economía de California. Dado que el objetivo del ETP era reducir los pagos de Seguro de Desempleo y promover el desarrollo económico, los evaluadores examinaron el impacto del ETP en los ahorros e ingresos del Seguro de Desempleo. El impacto del ETP es la diferencia entre lo que ocurrió con la presencia de los programas de capacitación del ETP y lo que hubiera ocurrido sin ellos. Estimaron que los 73 millones de dólares invertidos en el ETP dieron como resultado un ahorro de 63 millones de dólares en el Seguro de Desempleo y generaron 200 millones de dólares de ingresos adicionales. Combinando estos efectos y agregando 150 millones de dólares por efectos indirectos el impacto total del ETP es de más de 400 millones de dólares. Sin embargo, los autores advierten que la proporción de los impactos totales sobre la inversión inicial no debe ser interpretada como una tasa de retorno del programa, dado que no incluyeron todos los costos y rendimientos en sus cálculos. Por ejemplo, no incluyeron todos los costos de la capacitación cubiertos por los capacitandos individuales y por las compañías ni los costos administrativos del ETP. También excluyeron beneficios alcanzados después del primer año posterior a la capacitación, e ignoraron los beneficios para las compañías más allá de lo que se alcanzó para los empleados.

Utilizando el enfoque de la empresa representativa

El enfoque de la empresa representativa utiliza una metodología diferente al considerar la importancia relativa de la capacitación adecuada a las necesidades de una empresa financiada por el estado. Este enfoque no evalúa el comportamiento real de las empresas, sino que examina los efectos de la capacitación en los ingresos corrientes de las empresas representativas. El enfoque de la empresa representativa ha sido utilizado por diversos investigadores, incluyendo a Fisher y Peters (1998) y en estudios de importantes empresas consultoras, para analizar las ventajas relativas de la ubicación de los negocios. Esta metodología se basa en las declaraciones de ingresos y otros indicadores financieros de las empresas representativas en las industrias seleccionadas. El análisis toma en consideración los beneficios para las empresas, generalmente medidos con respecto al valor presente neto del ingreso futuro de la empresa (VPN) o la tasa interna de retorno (TIR) cuando se incluye el valor monetario de los diferentes incentivos y subsidios económicos de desarrollo. Así, una empresa que disfruta de deducciones en los impuestos de propiedad al ubicarse en un área específica mostrará un ahorro en los costos en comparación con una empresa idéntica ubicada en otra área en la que no se ofrece este incentivo.

Las actividades de desarrollo de la mano de obra se consideran en este análisis al incluir sólo el ahorro en costos por contar con el patrocinio gubernamental para la capacitación en lugar de que la empresa incurra en dicho costo. Donde puede aplicarse, también se incluye el reembolso estatal por los

salarios perdidos mientras el trabajador está capacitándose. Sólo son incluidos los programas que proporcionan capacitación específicamente a las empresas; los programas federales que se enfocan en beneficiar a los individuos son excluidos. No se hace intento alguno por estimar el incremento en la productividad del trabajador y su efecto en la productividad total de la empresa como resultado de la capacitación.

Fisher y Peters (1998) proporcionan las estimaciones más explícitas de las diversas actividades de desarrollo económico utilizando este método. Ellos hallaron que, en promedio, los incentivos no fiscales son el principal componente del total del paquete estatal de incentivos. En algunos casos, hasta el 90 por ciento de los incentivos estatales se derivaron de programas no fiscales. De los tres incentivos no fiscales (capacitación de la mano de obra, infraestructura y apoyos gubernamentales en general), los incentivos a la capacitación de la mano de obra son más importantes que los incentivos a la infraestructura para empresa de cualquier tamaño y actividad productiva o tipo de industria. Los incentivos a la capacitación son los incentivos no fiscales para las grandes empresas, definidas por el número de empleados.²⁰ No es sorprendente, que para las pequeñas empresas la capacitación no sea tan importante como los apoyos gubernamentales en general (los cuales incluyen una variedad de préstamos, garantías para préstamos y programas de subsidio principalmente para PyMEs). En conjunto, Fisher y Peters estiman que el paquete promedio de incentivos financieros no fiscales, expresados en términos de una equivalencia salarial a valor presente, fue de nueve centavos por empleado por hora. Esto significa que una planta típica puede absorber salarios más altos por hasta nueve centavos en una hora por empleado durante un periodo de 20 años sin afectar su balance de resultados. Para algunas empresas, la capacitación de la mano de obra está por encima de la mitad de los ahorros en la nómina. Agregando a este análisis los efectos del alza en productividad encontrados en otros estudios incrementaría la ventaja financiera de las empresas que reciben apoyos de estos programas.

Evidencia de la efectividad de los programas de capacitación financiados estatalmente

Obviamente, sin un número suficiente de estudios confiables de programas de capacitación para el trabajo financiados estatalmente en que basarse la evidencia de su efectividad es inconsistente. El pequeño grupo de evaluaciones de los programas estatales de capacitación generan las siguientes conclusiones. Holzer, *et al.* (1993) demuestra que los subsidios a la capacitación privada incrementan la capacitación y mejoran la productividad. Juntar los componentes esenciales que vinculan la capacitación con la productividad de la empresa y utilizar estimaciones del sector privado sugiere que la capacitación puede incrementar las ventajas en los costos de las empresas al aumentar la productividad más de

lo que se elevaron los salarios en una empresa, bajando así los costos unitarios (Barron, Berger y Black, 1997). Sin embargo, menores costos operativos no son suficientes para incrementar el empleo en una zona. También debe demostrarse que las empresas responden suficientemente a menores costos contratando a trabajadores adicionales o trasladándose hacia una zona que tenga esta ventaja de costo. La respuesta de las empresas a estas condiciones depende de diversos factores, incluyendo la disponibilidad de otros factores y el clima empresarial de la región en general. Moore *et al.* (por publicarse) encuentran evidencia de que los programas estatales de capacitación incrementan la productividad tanto de las empresas como de los trabajadores, aumentan los salarios, mejoran el ambiente laboral y amplían la mano de obra. Estos resultados se basan en un conjunto de estudios de programas estatales específicos de capacitación. Se requieren estudios de programas estatales de capacitación adicionales antes de que la solidez de estos resultados pueda contar con el mismo nivel de confianza con que cuentan los resultados de las evaluaciones de los programas federales de capacitación.

Conclusión

Las evaluaciones se han convertido en una parte integral de los programas de capacitación para el trabajo en el ámbito federal, pero en menor grado en los niveles estatales y locales. Las evaluaciones de programas federales han demostrado ser útiles a los funcionarios encargados de elaborar las políticas para determinar la reautorización y diseño de los programas de capacitación. Las pocas evaluaciones efectuadas de los programas estatales de entrenamiento también han producido resultados interesantes y útiles.

El contraste entre las evaluaciones en el ámbito federal y en el ámbito estatal y local ofrece intuiciones sobre la motivación para efectuar las evaluaciones y utilizar sus resultados. El número relativamente reducido de evaluaciones locales y estatales de programas de capacitación, u otros programas sociales, puede ser explicado por tres factores: 1) expectativas con respecto a la validez y utilidad de los resultados de las evaluaciones, 2) apoyo y voluntad política, 3) restricción de recursos. Primero, los programas sociales que están orientados a afectar las economías locales y regionales son difíciles de evaluar. Su alcance es más amplio que el resultado para los individuos y la escala de su intervención es típicamente muy pequeña para hacer una diferencia significativa en la región. Por lo tanto, las expectativas con respecto a la utilidad de evaluar dichos programas generalmente no compensa el costo de administrarla.

Segundo, los programas estatales y locales normalmente carecen de apoyo para las evaluaciones. Dado lo impreciso de las estimaciones y la probabilidad de que los resultados mostrarán muy poco —o nada— sobre los impactos del

programa, los grupos interesados en el programa frecuentemente son renuentes a presionar para realizar una evaluación. Los contribuyentes y los organismos de control del gobierno tienden a prestar más atención a errores en la contabilidad financiera que a la falta de efectividad de un programa. Los funcionarios de los gobiernos estatales y gubernamentales a cargo de programas de desarrollo económico se enfrentan con la realidad política y económica de que todos están en competencia con otras regiones para retener y atraer a los negocios. Para ellos, el asunto de realizar o no una actividad de desarrollo económico está relacionado con la competencia con otras jurisdicciones más que con la efectividad del programa. Estos funcionarios sostienen con frecuencia la visión de que si una empresa espera recibir incentivos para el desarrollo económico, uno de los cuales es la capacitación para el trabajo adaptada a sus necesidades, entonces los gobiernos estatales y locales deben proveer los incentivos apropiados si esperan competir. Así, en este ambiente competitivo, la voluntad política se conduce más por las percepciones y las expectativas que por los efectos en sí.

En contraste, el gobierno federal no está sujeto a la misma presión de los observadores que los gobiernos estatales y locales, y es seguro que las evaluaciones de sus programas rendirán generalmente resultados más precisos y en consecuencia más útiles. Además, las evaluaciones proporcionan medios por los cuales los políticos de los partidos de oposición pueden entablar compromisos para las subsiguiente legislación social.

Tercero, los recursos pueden llegar a ser también un factor, especialmente para los estados y las jurisdicciones locales más pequeños. Los recursos incluyen no sólo los gastos fuera del presupuesto para efectuar las evaluaciones, sino también suficiente habilidad y talento por parte del personal de estas jurisdicciones para que puedan entender y promover el uso de las evaluaciones. El costo de evaluar un programa no debe ser el principal obstáculo para llevar a cabo las evaluaciones excepto en las jurisdicciones más pequeñas. Normalmente, se gasta tanto o más dinero en auditar los estados financieros de un programa que en lo que costaría evaluar la efectividad del propio programa. Aun aquéllos interesados en asegurar que los fondos gubernamentales sean gastados en los fines para los que fueron asignados, deberían estar interesados en la efectividad del programa como lo están en sus transacciones financieras. De esta manera, se consideran importantes recursos en asegurar que el personal y los funcionarios encargados de diseñar las políticas poseen la capacitación y la información relevante para apreciar el valor de las evaluaciones y para entender y utilizar sus resultados.

No obstante, debe notarse que las evaluaciones fueron una parte integral de muchas iniciativas estatales y locales de reforma para la asistencia social durante los noventa. La diferencia entre aquellos programas y muchos de los programas de capacitación para el trabajo financiados estatalmente es la participación

federal. En muchos casos, con la reforma a la asistencia social, se le concedió a los estados dispensas para cambiar varios aspectos del programa federal sólo si aceptaban evaluar el programa. Mucho de lo que nosotros conocemos sobre la efectividad de diferentes experimentos con la reforma a la asistencia social proviene directamente de estas evaluaciones.

Por lo tanto, de estos ejemplos de evaluaciones de programas federales de capacitación para el trabajo, parece que las evaluaciones son instrumentos valiosos para configurar la política social. También es evidente que niveles más altos del gobierno, preferentemente del gobierno federal, requieren tomar el liderazgo al promover el uso de las evaluaciones de los programas sociales. Ellos cuentan con los recursos y las habilidades, y no están sujetos a presiones que disminuyan la voluntad política de efectuarlas. Aún más, las jurisdicciones estatales y locales pueden interesarse más en llevar a cabo las evaluaciones si se puede demostrar que las evaluaciones son creíbles y que ellas pueden ayudar a la mejora continua de sus programas. Con presupuestos gubernamentales ajustados y la impaciencia de los contribuyentes por el uso ineficiente de los dólares entregados al fisco, los gobiernos deberían recibir gratamente las oportunidades de volverse más rentables. Las evaluaciones proveen información valiosa para guiar las políticas y para diseñar programas acordes con las necesidades de aquellos a quienes pretenden servir.

Notas

1. Estas comparaciones se abstraen del Crédito Fiscal sobre la Renta Obtenido (EITC, por sus siglas en inglés) que se paga en los EU a los trabajadores de bajos ingresos con menores dependientes. En años recientes el EITC, que esencialmente es un subsidio dirigido a los salarios, totalizó cerca de 30 mil millones de dólares o aproximadamente igual al gasto total para las LMPs listadas en el texto.
2. Cuando se aplica el estimador pre-post del impacto neto, la situación pre-programa del grupo de participantes constituye la situación contrafactual.
3. Aunque las estimaciones de impacto neto apoyadas en diseños experimentales clásicos con muestreo aleatorio son consideradas el “estándar dorado” para las evaluaciones tienen limitaciones en el alcance del análisis. Friedman, Greenberg y Robins (1997) señalan que las evaluaciones de impacto neto sólo indican si un programa en particular funciona en promedio para una muestra bajo un marco de circunstancias definidas. Este enfoque no abre la “caja negra” para revelar qué aspectos del programa contribuyeron a su éxito o fracaso.
4. Dicho tipo de ejercicio puede ser realizado usando mediciones estandarizadas, como la distancia Mahalanobis (O’Leary, Nesporova y Samorodov 2001, p. 143).
5. Para ejemplos de programas de empleo evaluados usando un experimento de campo de diseños clásico véase Decaer y O’Leary (1995).
6. En este informe, dado que la principal variable dependiente de interés —colocación en un trabajo normal— es binaria, el modelo de regresión predice la probabilidad de recolocación. La estimación de mínimos cuadrados ordinarios (MCO) es un modelo de probabilidad lineal que puede producir estimadores sesgados. Los estimadores de

MCO pueden ser sesgados debido a que el rango de variación de la variable dependiente se restringe al intervalo unitario (0 a 1). Dado que las probabilidades de recolocación para los grupos de entrenamiento para el trabajo y de comparación oscilan generalmente entre el 40 y 60 por ciento, no es verosímil que el limitado rango de la variable dependiente sea una fuente de sesgo severo al estimar parámetros mediante MCO.

7. En esta aplicación, el modelo regresión es una declaración de una metodología de análisis de covarianza, donde X_1 a X_n son los covariados. Mohr (1992, pp. 83-87) trabaja sobre la extensión de un modelo de regresión sobre impacto del programa para incluir variables de control.
8. Un método para la creación de un grupo de comparación por emparejamiento basado en las características observables es explicado en el Anexo 4, Sección A.4.2*.
* N.T. No se incluye el anexo en el material publicado.
9. Heckman y Smith (1996, pp. 74-78) explican el método de la diferencia en las diferencias como un estimador de efecto fijo.
10. Un enfoque no paramétrico que también ajusta por los factores no observados es usado por Puhani (2002), quien proporciona un ejemplo para la creación de un grupo de comparación emparejado usando la probabilidad de participación en el programa (Participation Propensity Scores) que implícitamente toman en cuenta los factores no observables, así como los observables. Los estimadores de impacto son la diferencia simple de las medias entre ambos grupos. Heckman y Smith (1996, pp. 72-74) explicar el atractivo teórico del enfoque por probabilidad de participación (Propensity Score).
11. Los detalles técnicos sobre las propiedades de los estimadores: insesgado, eficiencia y robustez son explicados en el Anexo 5 de este manual*.
* N.T. No se incluye el anexo en el material publicado.
12. Estos son los errores estándar de las estimaciones puntuales de los impactos netos del programa.
13. Una revisión de este problema es elaborada por Greene (1993, pp. 273-277).
14. www.mathematica-mpr.com/3rdLevel/earlyimpact.htm.
15. Un relato corto de esta evaluación puede ser encontrado en *Employment Research*, W. E. Upjohn Institute for Employment Research, octubre 2001, vol. 9, núm. 4 en www.upjohn.org.
16. Recientemente, se han producido varias encuestas valiosas encaminadas a responder la pregunta sobre que tipo de entrenamiento para el trabajo funciona mejor para grupos demográficos en particular [King (2002), Barnow y King (2000), Heckman, LaLonde y Smith (1999)].
17. Oficina General de Contabilidad de los EU, junio, 1989, p. 33.
18. Obviamente existen muchos otros programas en operación de los gobiernos estatales y locales y de organizaciones no gubernamentales, algunos de los cuales han sido evaluados. Muchos de estos programas son descritos en Giloth (1998) y Bartik (2001). Nosotros nos enfocamos a los programas de entrenamiento para el trabajo con financiamiento estatal dirigidos a las empresas, con los propósitos de enfocarnos en programas que tienen la intención de afectar las economías locales y de contrastar éstos con los programas federales de entrenamiento para el trabajo dirigidos a los trabajadores como individuos.
19. Por ejemplo, Walter y Greenstreet (1989) hallaron que las decisiones de selección de ubicación son sensibles a los programas de entrenamiento para el trabajo.

20. Tres categorías de tamaño de planta fueron utilizadas. Se clasificó como grandes plantas a aquellas con empleados por arriba del 75 percentil de sus respectivas industrias. Las pequeñas plantas son aquellas cuyo número de empleados estuvo por debajo del 25 percentil. Por lo tanto, no se dio una delineación absoluta del tamaño de la firma debido a que el tamaño de la firma varió significativamente entre las industrias.

Referencias bibliográficas

- ASHENFELTER, Orley (1978), "Estimating the Effect of Training Programs on Earnings", *Review of Economics and Statistics*, 6(1): 47-57.
- BARNOW, Burt S. (1989), "Government Training as a Means of Reducing Unemployment" en *Rethinking Employment Policy*, D. Lee Bawden and Felicity Skidmore, eds. Washington, DC: Urban Institute.
- BARNOW, Burt S. y Christopher A. KING (eds.) (2000), *Improving the Odds: Increasing the Effectiveness of Publicly Funded Training*. Washington, DC: The Urban Institute Press.
- BARNOW, Burt S. y Jeffrey A. SMITH (2002), "Performance Management of US Job Training Programs", presentado en la conferencia *Job Training in the United States: History, Effectiveness and Prospects*, Augusta, Michigan: W.E. Upjohn Institute for Employment Research.
- BARRON, John M., Mark C. BERGER y Dan A. BLACK (1997), *On-the-job Training*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- BARTIK, Timothy J. (1991), *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- BARTIK, Timothy J. (2001), *Job for the Poor: Can Labor Demand Policies Help?* Nueva York, NY: Russell Sage Foundation and MI: W.E. Upjohn Institute for Employment Research.
- BURGHARDT, John, Peter Z. SCHOCHET, Sheena MCCONNELL, Terry JOHNSON, R. Mark GRITZ, Steven GLAZERMAN, John HOMRIGHAUSEN y Russell JACKSON (2001), *Does Job Corps Work? Summary of the National Job Corps Study*. Documento núm. PR01-50. Princeton, NJ: Mathematica Policy Research, Inc. (Junio).
- DECKER, Paul T. y Christopher J. O'LEARY (1995), "Evaluating pooled evidence from the re-employment bonus experiments". *Journal of Human Resources*, vol. 30, núm. 3, pp. 534-550.
- DECKER, Paul T. e Irma PEREZ-JOHNSON (2002), "Individual Training Accounts and Eligible Provider Lists", presentado en la conferencia *Job Training in the United States: History, Effectiveness and Prospects*, Augusta, Michigan: W.E. Upjohn Institute for Employment Research.
- FISHER, Peter S. y Alan H. PETERS (1998), *Industrial Incentives: Competition among American States and Cities*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- FRIEDLANDER, Daniel, David H. GREENBERG y Philip K. ROBINS (1997), "Evaluating Government Training Programs for the Economically Disadvantaged", *Journal of Economic Literature* 35(4), diciembre.

- GILLOTH, Robert (1998), *Jobs and Economic Development: Strategies and Practice*. Thousand Oaks, CA: Sage Publications.
- GREENE, William H. (1993), *Econometric Analysis*. Nueva York: Macmillan.
- HECKMAN, James J. (1979), "Simple Selection Bias as a Specification Error". *Econometrica* 47(1): 153-161.
- HECKMAN, James J., Robert J. LALONDE y Jeffrey A. SMITH (1999), "The Economics and Econometrics of Active Labor Market Programs", en el *Handbook of Labor Economics*, volumen #A, Orley Ashenfelter y David Card, editores. Amsterdam: Elsevier.
- HECKMAN, James J., REBECCA Roselius y Jeffrey SMITH (1994), "US Education and Training Policy: A Re-evaluation of the Underlying Assumptions Behind the 'New Consensus'". En *Labor Markets, Employment Policy and Job Creation*, Lewis Solomon y Alec Levenson, eds. Boulder, CO: Westview Press.
- HECKMAN, James J. y Jeffrey A. SMITH (1996), "Experimental and Nonexperimental Evaluations", en Schmid, O'Reilly y Schomann, eds. *International Handbook of Labour Market Policy and Evaluation*. Cheltenham, RU: Edward Elgar, pp. 37-88.
- HOLLENBECK, Kevin M. (2002), "Washington's Workforce Development System Pays Off". *Employment Research*. 9(4): 4-6. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- HOLZER, Harry J., Richard N. BLOCK, Marcus CHEATHAM y Jack H. KNOTT (1993), "Are Training Subsidies for Firms Effective? The Michigan Experience", *Industrial and Labor Relations Review*, vol. 46(4), julio, 635-636.
- KING, Christopher A. (2002), "The Effectiveness of Publicly Financed Training in the US", presentado en la conferencia *Job Training in the United States: History, Effectiveness and Prospects*, Augusta, Michigan: W.E. Upjohn Institute for Employment Research.
- KLUGE, Jochen y Christoph M. SCHMIDT (2002), "Can Training and Employment Subsidies Combat European Unemployment?" *Economic Policy: A European Forum*.
- LEIGH, Duane E. (1990), *Does Training Work for Displaced Workers?: A Survey of Existing Evidence*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- MADDALA, G.S.(1983), *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.
- MALLAR, Charles, Stuart KERACHSKY, Craig THORNTON y David LONG (1980), *An Evaluation of the Economic Impact of the Job Corps Program*. Proyecto de informe 80-60. Princeton, NJ: Mathematica Policy Research, Inc.
- MANGUM, Garth L. (1968), *MDTA: Foundation of Federal Manpower Policy*. Baltimore, MD: The Johns Hopkins Press.
- MOHR, Lawrence B. (1992), *Impact Analysis for Program Evaluation*. Londres: Sage.
- MOORE, Richard W., Daniel BLAKE, G. Michael PHILLIPS y Daniel MCCONAUGHY (por publicarse), *Training that Works: Lessons from California's Employment Training Panel Program*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- OCDE (2001), *OCDE Employment Outlook*. París: Organización para la Cooperación y el Desarrollo Económico (junio).
- O'LEARY, Christopher J., Alena NESPOROVA y Alexander SAMORODOV (2001), *Manual on Evaluation of Labour Market Policies in Transition Economies*. Genova: ILO.

- O'LEARY, Christopher J. y Robert A. STRAITS (2002), "Intergovernmental Relations in Employment Policy: The United States Experience", en *Labour Market Policy and Federalism: A Comparative Perspective*, Alain Noël, ed. Montreal y Kingston: McGill-Queen's University Press.
- ORR, Larry L., Howards S. BLOOM, Stephen H. BELL, Fred DOOLITTLE, Winston LIN y George CAVE (1996), *Does Training for the Disadvantaged Works?: Evidence from the National JTPA Study*. Un estudio de Abt Associates. Washington, DC: Urban Institute Press.
- PUHANI, Patrick A. (2002), "Advantage through Training in Poland?: A Microeconomic Evaluation of the Employment Effects of Training and Job Subsidy Programs". *Labour* 16(3): 569-608.
- SOCIAL POLICY RESEARCH ASSOCIATES (2001), *PY 99 SPIR Data Book*. Oakland, California: Social Policy Research Associates (abril).
- STROMSDORFER, Ernst W. y George FARKAS (eds) (1980), *Evaluating Studies Review Annual* — Volumen 5. Beverly Hills, CA: Sage.
- US GENERAL ACCOUNTING OFFICE (1989), "Job Training Partnership Act: Services and Outcomes for Participants with Different Needs", US Government Printing Office, junio, p. 33.
- WALKER, Robert y David GREENSTREET (1989), "Public Policy and Job Growth in Manufacturing: An Analysis of Incentives and Assistance Programs". Artículo presentado en la 36a. Reunión Norte Americana de la Regional Science Association, Santa Barbara, CA, noviembre 10-12.
- WANDNER, Stephen A., David E. BALDUCCHI y Amanda B. SPICKARD (2001), "Expenditures on Active and Passive Labor Market Policy in the United States Estimates for Fiscal Year 2001". Presentado en el taller internacional "Active Labor Market Programs: Improvement of Effectiveness" patrocinado por el Ministerio del Trabajo y Desarrollo Social de la Federación Rusa, Banco Mundial y Fundación Rusa para las Reformas Sociales, octubre 2-3, Moscú, Rusia.

Capítulo 12

Evaluación de políticas de desarrollo económico local: teoría y práctica*

por
Jeffrey Smith,
Facultad de Economía, Universidad de Maryland, Estados Unidos

* Este capítulo se ha beneficiado de las exposiciones o comentarios de Tim Bartik, Dan Black, Michael Lechner, Alistair Nolan, Miana Plesca, Elliot Stern y Alex Whalley, de la lectura de los capítulos de Tim Bartik y Randy Eberts y Chris O'Leary (que fueron escritos anterior a éste), y de las discusiones durante la conferencia de la OCDE en Viena en noviembre 2002.

1. Introducción

Las políticas y programas emprendidos por los gobiernos y las agencias privadas para incrementar el desarrollo económico local podría ser que tuvieran efectos tanto positivos como negativos. En algunos casos, la falta de efectividad podría deberse a que el programa tiene un diseño pobre, o un financiamiento inadecuado. En otros casos, la falta de efectividad podría ser resultado del programa que existe para canalizar dinero a entidades, individuos o grupos políticamente influyentes, bajo la justificación del desarrollo económico local. Cuando los programas no producen beneficios en términos de desarrollo económico local, detectarlo permite que los fondos escasos fluyan hacia otras actividades más benéficas, o que regresen hacia el siempre atribulado contribuyente. Cuando los programas sí producen beneficios, detectarlo puede generar apoyo político para la continuidad de dicho programa o incluso para su expansión.

La evidencia sobre la eficacia de las políticas y programas de desarrollo económico local provienen de las evaluaciones. Este capítulo presenta un panorama de la bibliografía actual sobre cómo evaluar los programas. Sabiamente la bibliografía sobre la evaluación de programas ha avanzado rápidamente a lo largo de los últimos quince años. Por ejemplo, los principales avances en relación con la “efectividad del tratamiento heterogéneo” —diferentes impactos de un programa para diferentes personas, entidades, condados, ciudades o grupos afectados por una política— tienen repercusiones en la práctica de las evaluaciones y en cómo pensar acerca del diseño de la evaluación e interpretación. De igual manera, los avances técnicos importantes en los métodos no-paramétricos y semi-paramétricos permiten el uso más flexible de los datos disponibles, pero al mismo tiempo se crea la demanda por datos de alta calidad que son requeridos por dichos métodos para producir estimaciones confiables. Los experimentos sociales ahora son una rutina (al menos en los Estados Unidos) en áreas como la evaluación de los programas públicos de empleo y capacitación. La práctica de evaluaciones en su mayoría, desafortunadamente permanecen ancladas en los setenta. Uno de los principales objetivos de este capítulo es proporcionar, relativamente, una guía práctica, no técnica, de estos avances.

Este capítulo aborda algunos de los mismos temas que los capítulos de Tim Bartik (2004) y Randy Eberts, y de Christ O’Lear (2004), pero con suficientes diferencias para hacerlo un complemento, en lugar de un sustituto, para aquellos capítulos. En particular existen cinco diferencias que merecen resaltarse. Primero, este capítulo dedica mucha mayor atención a los diferentes estimadores de evaluación econométrica en la bibliografía, y provee una cantidad abundante de indicadores en la creciente bibliografía en este tema. Un tema clave del capítulo es la elección de un estimador apropiado dado los datos disponibles, el ambiente económico y las características institucionales del programa que está siendo evaluado. Segundo, este capítulo dedica mucha mayor atención a la bibliografía emergente sobre los efectos del tratamiento heterogéneo, y cómo dichos efectos influyen en el diseño e interpretación de la evaluación. Tercero, este capítulo se preocupa más por las implicaciones de efectividad del equilibrio general para la evaluación de políticas. Cuarto, este capítulo enfatiza que el efectuar una evaluación no podría tener sentido en todos los casos, en particular para los programas más pequeños. El tiempo invertido en la lectura de buenas evaluaciones en programas similares puede rendir resultados más útiles que una evaluación ineficaz basada en datos deficientes realizada por un evaluador poco calificado que usó métodos inapropiados. Finalmente, la perspectiva implícita en este capítulo es que la evaluación, tomada seriamente, representa un método para asegurar que los administradores del programa fomenten las vayan más allá de los objetivos de sus protagonistas —entiéndase contribuyentes fiscales y donadores— en lugar de simplemente transferir los recursos a las partes interesadas, como son los operadores de los programas, entidades políticamente favorecidas, o ellos mismos. En la práctica, existen muchas evaluaciones de baja calidad con el propósito de ocultar dicho comportamiento; precisamente por esta razón es importante ser muy claros sobre lo que constituye una buena evaluación y diseñar instituciones que reduzcan el flujo de falsas apariencias, con evaluaciones de baja calidad.

El resto del capítulo está organizado de la siguiente manera: La sección 2 describe los problemas de evaluación y trata parámetros de interés. La sección 3 proporciona un panorama de la teoría para la evaluación econométrica del programa a un nivel relativamente no-técnico, con suficientes sugerencias hacia la bibliografía. La sección 4 revisa las dos posibilidades serias para la evaluación econométrica del programa: la autoevaluación de los participantes y los indicadores de desempeño administrativos. La sección 5 expone la práctica de la evaluación, en el sentido amplio de alternativas que enfrenta una organización que considera emprender una evaluación, cómo determinar si vale la pena o no el realizar una evaluación, quién debe realizar la evaluación, y cómo asegurarse que ésta sea buena. La sección 6 concluye y reafirma los temas principales en este capítulo.

2. Programas y parámetros

2.1. Tipos de programas de desarrollo económico local

Los programas de desarrollo económico local incluyen una amplia gama de iniciativas, desde programas diseñados para mejorar el capital humano de trabajadores individuales hasta subsidios financieros y en especie para equipos de atletismo profesionales, y subsidios fiscales diseñados para atraer cierto tipo de negocios para zonas empresariales, y otros. Bartik (2004) presenta una buena lista en su Tabla 12.1 a que incluye un conjunto de actividades más estrecho, que el que tengo aquí en mente; Bartik (2003a) describe estas políticas en mayor detalle.

Para este capítulo, dos dimensiones de dichos programas tienen particular relevancia, ya que influyen en las elecciones sobre la recolección de datos y métodos de evaluación, como describo en detalle a continuación. La primera dimensión corresponde a las unidades directamente tratadas por la intervención. Dependiendo del programa, éstas pueden ser trabajadores individuales, algunas o todas las empresas en un área, ciudades, pueblos o distritos, estados y países completos. La segunda dimensión, consiste en las unidades que la teoría sugiere que afectarán al programa. En algunos casos, en especial con los programas que no se espera que tengan muchos efectos externos, ambas dimensiones podrían coincidir. Por ejemplo, los programas de capital humano a pequeña escala podrían tener poca efectividad en individuos a diferencia de aquellos que reciben el capital humano adicional. En otros casos, las dos dimensiones no coinciden. Por ejemplo, un programa podría tener efectos positivos en las unidades tratadas y efectos negativos en las unidades no tratadas, como cuando se subsidia a una clase de empresas pero no a sus competidores. Y aun en otros casos, los programas pueden producir un efecto positivo del excedente de población de los centros urbanos, como cuando un nuevo parque atrae negocios y residentes a un área, e incrementa el valor de la propiedad en los vecindarios a su alrededor.

2.2. Anotación

Las discusiones más entendidas sobre las políticas de los programas de desarrollo económico con frecuencia se enfocan en sus “efectos”, como si los “efectos” de un programa representasen una única y bien definida entidad. Los programas que tienen una variedad de efectos representan un importante tema en este capítulo. En la bibliografía académica, esta discusión cae bajo el encabezado de efectos heterogéneos del tratamiento. La aborda como la concepción de los efectos de un programa cambia y se amplía cuando consideramos que un programa puede tener diferentes efectos en cada unidad en que participa y, en algunos casos, incluso en unidades que no participan.

Para que ésto sea más claro, les presento algunas anotaciones muy simples que servirán para aclarar que existen significados precisos a lo largo de este capítulo. Sin embargo, el capítulo está escrito de tal manera que no requiere una clara comprensión de la anotación para entender el significado; los lectores que sean enemigos acérrimos de las anotaciones pueden simplemente leerla por encima.¹

Deje que la Y denote un resultado variable. Para un individuo, ésto puede significar ingresos, empleo o estado de salud. Para una empresa puede significar ganancias, ventas o empleo. Para una localidad puede significar población, o alguna medición de la calidad de aire o del crecimiento económico. Ahora imaginemos dos mundos para cada unidad, un mundo en que la unidad participa en el programa bajo estudio y otro en el que no. Nosotros podemos imaginar el valor de la unidad Y en cada uno de estos mundos, y denotar el valor en el mundo donde la unidad participa como Y_{1i} , y el valor en el mundo donde la unidad no participa como Y_{0i} , y es donde la “ i ” se refiere a una unidad en particular.

2.3. Parámetros de interés

Usando esta anotación, el efecto de un programa sobre una unidad “ i ” está dado por

$$\Delta_i = Y_{1i} - Y_{0i}$$

En otras palabras, la bibliografía define el efecto de un programa en una unidad “ i ” como la diferencia en resultados entre un mundo donde esa unidad participa y otra unidad donde no lo hace. Es aquí en donde el problema de la evaluación consiste en estimar cualquiera de los dos resultados que no se observan en los datos.

Muchos de los diversos parámetros de interés definidos y examinados en la bibliografía sobre evaluación de programas son promedios de impacto específico de unidades (Δ_i) sobre diversos conjuntos de unidades políticas relevantes. El parámetro de interés más común es el Efecto de Tratamiento Promedio sobre las Tratadas (ATET, por sus siglas en inglés), o simplemente “tratamiento sobre las tratadas” para abreviar. Este parámetro indica el efecto promedio del programa con los participantes actuales. En términos de la anotación que se acaba de definir, es igual a

$$\Delta^{TT} = E(D_i | D_i=1) = E(Y_{1i} - Y_{0i} | D_i=1)$$

donde D_i es una variable dicotómica para los participantes actuales, de modo que $D_i=1$ para las unidades que participan en el programa y $D_i=0$ en cualquier otro caso, y E denota el operador de esperanza, donde la esperanza está condicionada a la propiedad a la derecha de la barra vertical (“|”). Una estimación de la ATET,

combinada con una estimación del costo promedio del programa por unidad participante, permite un análisis del costo beneficio a la pregunta de si se debe conservar o desechar un programa existente.

El Efecto de Tratamiento Promedio (ATE, por sus siglas en inglés) representa un segundo parámetro de interés potencial. Este parámetro promedia el efecto de tratamiento sobre todas las unidades en una población, incluyendo tanto a participantes como a no participantes. En términos de la anotación,

$$\Delta^{\text{ATE}} = E(Y_{1t} - Y_{0t})$$

El ATE responde preguntas sobre la política relacionada a los programas universales —programas donde cada unidad en poblaciones bien definidas participan. Al momento de considerar hacer que un programa voluntario sea obligatorio, los diseñadores de políticas requieren de estimaciones precisas del ATET y del ATE, las cuales pueden tener fuertes diferencias si aquellas unidades que eligieron no participar en un programa voluntario lo hacen porque sólo tendrían una efectividad reducida o incluso diferencias negativas. Un ejemplo aquí sería tomar un programa voluntario de asistencia en la búsqueda de empleo o de entrenamiento para un puesto, para trabajadores desplazados, y hacerlo obligatorio (por ejemplo, a través de requerir de la participación a manera de recibir los beneficios completos del seguro de desempleo).

Una tercera categoría de parámetros consiste en los Efectos Marginales del Tratamiento Promedio, o MATEs (por sus siglas en inglés). El Efecto Marginal de Tratamiento Promedio mide el efecto promedio de un programa entre un grupo con algún margen relevante. Por ejemplo, supongamos que el programa bajo observación en este momento atiende a empresas con menos de 20 empleados, y la propuesta bajo consideración consiste en una expansión para incluir a empresas de 21 a 30 empleados. En esta situación, el parámetro de interés consiste en el efecto promedio de participación de empresas con 21 a 30 empleados. Este parámetro puede diferir del ATET, el cual proporcionaría el efecto promedio sobre las empresas existentes que participan (aquellas con 1 a 20 empleados), y podría ser mayor o menor dependiendo de la naturaleza del tratamiento del programa y su relación con el tamaño de la empresa. Comparar un MATE con el correspondiente costo marginal de expansión (o contratar) un programa arroja un análisis costo-beneficio para la expansión o reducción del programa. Observe que un MATE diferente aplicado a cada margen, expandiendo un programa para incluir un conjunto de unidades, podría dar diferentes resultados al expandirlo para incluir otro conjunto de unidades. Una categoría final de parámetros, llamada Efectos Locales del Tratamiento Promedio, o LATEs (por sus siglas en inglés), está desarrollada en la Sección 3.4.

Los parámetros presentados hasta ahora pueden o no captar la efectividad del equilibrio general, dependiendo del diseño del análisis. La efectividad general

del equilibrio se establece en relación con la efectividad inmediata sobre las unidades tratadas, y el resultado de cambios en el comportamiento de las unidades no tratadas en respuesta al programa. Dichos cambios pueden ocurrir directamente (una empresa con 11 empleados despidió a uno con el propósito de volverse elegible para un programa que sirve a empresas con 10 o menos empleados) o indirectamente a través de cambios en los precios, como cuando un subsidio al pago de colegiaturas incrementa la oferta de trabajadores calificados y por lo tanto les baja el salario. Consideren un programa en el ámbito estatal que subsidia la capacitación a una clase específica de empresas. Algunos estados tienen este tipo de programa y otros no. Una estimación del ATET sobre las empresas que reciben el subsidio capturarán únicamente los efectos directos en el empleo, la productividad, las ventas y otros, para dichas empresas. Por el contrario, una estimación del ATET en los estados que adoptan el subsidio capturarán cualquier efecto de equilibrio general en el nivel estatal, incluyendo reducciones en el empleo en las empresas no subsidiadas, pero no efectos de equilibrio general que operen a través de los límites estatales.

El traer los efectos del equilibrio general al marco agrega algunos parámetros adicionales de interés. Por ejemplo, ahora podríamos tener un interés en lo que la bibliografía denomina como Efecto del Tratamiento Promedio en las no-Tratadas (ATNT, por sus siglas en inglés). Este parámetro mide el efecto promedio del programa en las unidades que no participan en él, ya sea por elección o porque no cumplen los criterios de elegibilidad. Para ejemplificar esto, considere el caso de un programa que proporciona apoyo en la búsqueda de trabajo a un grupo particular de trabajadores. Ahora, estos trabajadores buscan más y de manera más inteligente que antes, y esperamos que encuentren trabajos de manera más rápida. Pero ¿qué sucede con los otros en el mercado laboral? Primero, algunos trabajos que hubieran sido ocupados por otros, ahora son ocupados por los individuos que recibieron el apoyo para la búsqueda de trabajo. Como resultado, los no tratados encuentran trabajo de manera más lenta. Segundo, también para ellos puede tener sentido modificar su intensidad en la búsqueda de trabajo. Ambos factores llevan a un programa, que proporciona servicios a un grupo para tener efectos en otros grupos —son efectos que importan al ponderar el valor de un programa. Calmfors (1994) proporciona una útil (y relativamente accesible) introducción a los aspectos del equilibrio general en el contexto de las políticas activas para el mercado laboral.

3. Teoría

Esta sección proporciona una breve introducción a cada una de las principales categorías de métodos de evaluación econométrica. Cada una de las Secciones 3.1 a 3.6 son consideradas una categoría de método, y la Sección 3.7 considera cómo elegir entre ellas. A pesar de que este capítulo presenta a los diversos estimadores

como si fueran platillos en un buffet, donde el evaluador puede escoger cual utilizar, basándose en un nombre llamativo o que estén asociados con personas famosas, o por ser el estimador que tengan del día; de hecho, la selección de los estimadores debidamente realizada, debe apegarse a reglas estrictas. Cada una de las categorías de estimadores examinadas en las siguientes subsecciones proporciona la respuesta correcta sólo bajo ciertas suposiciones. Un evaluador que escoge un estimador, debe considerar cuidadosamente la naturaleza de los datos disponibles, la naturaleza institucional del programa —particularmente como se da la participación— y los parámetros de interés. En algunos casos, esto constituye otro de mis temas: la falta de datos adecuados puede provocar que ningún estimador pueda proveer una respuesta correcta, en cuyo caso el evaluador debe simplemente parar y reportar el hecho.

Para los lectores interesados en aprender más, la bibliografía proporciona otras encuestas sobre todo o parte de este material, abarca desde lo menos técnico hasta lo más técnico. En el extremo menos técnico, ver Moffitt (1991, 2003), Winship y Morgan (1999), Smith (2000), Ravallion (2001), y Smith y Sweetman (2001). A un nivel moderadamente técnico, ver Angrist y Krueger (1999), Blundell y Costa Dias (2000, 2002), y Heckman, LaLonde y Smith (1999), excepto para la Sección 7. Para las presentaciones verdaderamente técnicas ver la sección 7 de Heckman, LaLonde y Smith (1999), y Heckman y Vytlačil (2004). Algunos textos econométricos estándar incluyen presentaciones que enfatizan los aspectos enfocados en la bibliografía sobre evaluación. A este respecto ver Wooldridge (2002), para los estudiantes universitarios, y a Green (2002) o Wooldridge (2001) para los estudiantes con nivel de posgrado.

3.1. Experimentos sociales

Los experimentos sociales representan el utensilio más poderoso dentro de la caja de herramientas del evaluador, sin embargo no por ser la herramienta favorita es la más efectiva, por lo que los experimentos sociales sirven mejor al evaluador en algunos contextos que en otros. Para ver por qué a los evaluadores les agradan los experimentos sociales, considere un tratamiento sin efectos externos y suponga que buscamos determinar el impacto del tratamiento en aquellas ya tratadas. El principal problema en la investigación de las evaluaciones (casi) siempre consiste en la selección no aleatoria para el tratamiento. Debido a la selección no aleatoria para el tratamiento, uno no puede simplemente comparar los resultados de las unidades tratadas con los de las unidades no tratadas, para determinar el impacto de tratamiento. En los términos de la anotación antes definida, no podemos confiar en los resultados promedio de las unidades no tratadas, $E(Y_0 | D=0)$, para representar con exactitud los resultados que las unidades tratadas hubieran experimentado en caso de no haber sido tratadas, $E(Y_0 | D=1)$. El encontrar una buena aproximación a este contrafactual representa la parte difícil de la estimación de tratamiento en el parámetro de las tratadas (debido a los

resultados de las unidades participantes, $E(Y_1 | D=1)$, que aparecen directamente en los datos). El problema de selección no aleatoria en el tratamiento se denomina problema de sesgo de selección en la bibliografía de evaluación econométrica. Es muy importante distinguir el problema clásico de sesgo de selección sobre el resultado de seleccionar las unidades no tratadas, de la selección no aleatoria para el tratamiento basado en el efecto de tratamiento. Sólo recientemente, el último tipo de selección ha recibido atención sustancial en la bibliografía; este tipo de selección es el que hace, por ejemplo, el impacto medio del tratamiento en las tratadas, a diferencia del efecto del tratamiento promedio en programas que no tratan a todas las unidades elegibles.

Los experimentos sociales solucionan el problema de sesgo de selección a través de la construcción directa del contrafactual, usualmente no observado, sobre lo que las unidades participantes habrían experimentado en el caso de no haber participado. En particular en los experimentos sociales, las unidades que de otra manera hubieran recibido el tratamiento, son excluidas aleatoriamente de participar. Bajo ciertos supuestos, los resultados de estas unidades excluidas aleatoriamente proporcionan una estimación del contrafactual medio faltante, dado por $E(Y_0 | D=0)$. Esta capacidad para obtener el contrafactual bajo, lo que en muchos (pero no todos) los contextos, representa suposiciones bastante factibles y define el poder de los experimentos y explica su atractivo para los evaluadores.

Considerando que las virtudes de los experimentos son bastante conocidas y también extensamente detalladas en el capítulo de Bartík, yo mejor me enfoco en algunos de los aspectos conceptuales y limitaciones relacionadas con los experimentos. Esta exposición no pretende cubrir a aquellos que desean evitar la realización de los experimentos por que quieran mantener un aura de incertidumbre acerca de los impactos de los programas que “aman” (o se benefician financieramente de ellos, lo que frecuentemente se relaciona suma de la misma manera). En su lugar se trata de que los experimentos no sean usados cuando no corresponde, o no pueden responder a las preguntas de interés, y para asegurarse que sean interpretados correctamente cuando sean usados.

La primera limitación de los experimentos es que no pueden responder a todas las preguntas de interés. Esta limitación tiene tres facetas, que serán cubiertas a su debido momento. Primera, la aplicación aleatoria no es factible en muchos casos. La evidencia sugiere que la democracia aumenta el crecimiento económico, pero no se puede asignar aleatoriamente la democracia a los países. De manera similar, los factores políticos pueden prohibir la aplicación aleatoria de subsidios a las empresas, o el desarrollo de subvenciones aleatoriamente para las ciudades y los pueblos.

Segunda, los datos experimentales pueden o no capturar los efectos de equilibrio general de los programas. Ya sea que dependan o no de las unidades

afectadas por cualquier efecto de equilibrio y las unidades que son asignadas aleatoriamente en el experimento. Si existen efectos de desplazamiento hacia unidades no asignadas aleatoriamente, como cuando un programa para pequeñas empresas tiene efecto en empresas de mediano tamaño, estos efectos serán extrañados. Similarmente, los efectos positivos pasarán inadvertidos en una evaluación que sólo seleccione al azar las empresas a las tratadas, en lugar de seleccionar aleatoriamente al nivel de localidad.

Finalmente, los experimentos proporcionan la distribución de resultados experimentados por las unidades tratadas y la distribución de resultados experimentados por las no tratadas. Ellos no proporcionan la relación entre estas dos distribuciones; dicho de otra manera, los datos experimentales no indican si una unidad tratada que obtuvo un buen resultado también lo habría obtenido en caso de no recibir el tratamiento. En términos técnicos, este experimento proporciona los resultados marginales de distribuciones, pero no la distribución conjunta. Como consecuencia, sin mayores suposiciones no experimentales, los datos experimentales no identifican los parámetros que dependen de la distribución conjunta de los resultados, tal como la discrepancia de los impactos. Ver Heckman, Smith y Clements (1997) para una exposición extensa de estas cuestiones y de la diversidad de métodos para obtener estimaciones de estos parámetros.

La segunda limitación más importante de los experimentos es que en ciertas ocasiones las dificultades prácticas, asociadas con la implementación de los experimentos, pueden complicar su interpretación. Los lectores interesados en un tratamiento más general de la implementación de experimentos sociales deben consultar a Orr (1998), así como en la implementación de los reportes o resúmenes asociados con las principales evaluaciones experimentales, tales como Hollister, Kemper y Maynard (1984), Doolittle y Traeger (1990), Newhouse (1994) entre otros. La Revista de los Experimentos Sociales recopilada por Greenberg y Shroder (1997), presenta una lista comprensiva de todos los experimentos sociales, así como indicaciones sobre detalles acerca de su diseño, implementación y hallazgos.

En primer lugar, debido a que la evaluación experimental tiende a tener una mayor predisposición a efectos perjudicial en la operación de programas locales que una evaluación no experimental, los experimentos en sistemas descentralizados o federales (tales como los de Estados Unidos y Canadá) con frecuencia tienen problemas de validez externa, debido a la asignación no aleatoria de programas locales dentro del experimento. Esto fue un aspecto en el estudio nacional de la JTPA de los Estados Unidos, donde se tuvo acceso a 200 de los 600 centros locales de capacitación para encontrar 16 que estuvieran dispuestos a participar en la evaluación experimental. Además de tratar de mantener el diseño experimental relativamente discreto, y de ofrecer pagos compensatorios (cerca de 1 millón de dólares destinados en el estudio de JTPA),

poco puede hacerse acerca de ésto más que comparar las características de los programas locales participantes y no participantes, y evitar generalizaciones excesivamente ambiciosas acerca de los resultados.

Segundo, como se describe por ejemplo en Heckman y Smith (1995) los experimentos pueden sufrir de sesgo aleatorio. Esto sucede cuando los individuos se comportan diferente debido a la presencia de la aleatoriedad. Por ejemplo, si las unidades bajo estudio pueden emprender actividades que complementan el tratamiento antes de recibirlo, tienen menor incentivo de hacerlo durante un experimento, porque podrían ser excluidos aleatoriamente del tratamiento. Observe que el sesgo aleatorio difiere de los efectos Hawthorne. El último ocurre cuando los individuos al ser evaluados modifican su comportamiento en respuesta a ser observados, ya sea en el contexto de una evaluación experimental o de una no experimental. Existe poca evidencia empírica sobre la importancia del sesgo aleatorio.

Tercero, dependiendo en parte de la ubicación del muestreo aleatorio dentro del proceso por el cual las unidades llegan a recibir el tratamiento, el abandono desde el tratamiento de grupo podría causar problemas para la interpretación de las estimaciones experimentales de impacto. El abandono aquí se refiere a la salida del tratamiento después del muestreo aleatorio, quizás porque parece menos atractivo una vez que se conoce por completo. El muestreo aleatorio de las unidades al principio del proceso de participación tiende a incrementar el abandono. Según se detalla en Heckman, Smith y Taber (1998), el abandono es una característica usual de la evaluación experimental de las políticas activas en el mercado laboral. Las respuestas comunes toman dos formas. Bajo la primera, la interpretación de las estimaciones del impacto cambia y se convierte en el medio de impacto de la oferta del tratamiento, en lugar de recepción del tratamiento. Bajo la segunda, la estimación del impacto se ajusta usando el método de Bloom (1984). Ver Heckman, LaLonde y Smith (1999), Sección 5.2, para mayores detalles y explicación del origen del ajuste.

Cuarto, como se expone a detalle en Heckman, Hohmann, Smith y Khoo (2000), en algunos contextos, las unidades del grupo de control pueden recibir un tratamiento similar al ofrecido al grupo experimental del tratamiento de otras fuentes. Su análisis considera el caso de programas de empleo y capacitación y muestran que, al menos en el ambiente institucional descentralizado de los Estados Unidos (donde muchas agencias federales y estatales ofrecen capacitación subsidiada de diferentes tipos, la sustitución es bastante común. En este caso, los resultados del grupo de control no representan lo que las unidades tratadas hubieran experimentado en caso de no haber recibido el tratamiento. En su lugar, éstos representan alguna combinación de resultados de unidades no tratadas y tratadas alternativamente. La bibliografía señala tres respuestas al sesgo de sustitución. Como en el caso de abandonos, una consiste en la reinterpretación

del parámetro —esta vez como la medida de la diferencia estimada de la medida entre el tratamiento que se evalúa y lo que hubieran recibido las unidades tratadas si el tratamiento no hubiera estado disponible, lo que significa algunas veces no ser tratado y otras veces recibir otro tratamiento. La segunda respuesta consiste en ajustar la diferencia de medias estimadas, al dividirla por la diferencia en la fracción tratada entre los grupos de tratamiento y de control. Esto representa una generalización del estimador de Bloom (1984) y requiere que el tratamiento sustituto tenga un impacto similar al tratamiento que es evaluado. Finalmente, la tercera respuesta consiste en usar los datos experimentales para hacer una evaluación no experimental. Ver Heckman, Hohmann, Smith y Khoo (2000) para más detalles.

Como se ha indicado por ejemplo en Smith y Sweetman (2002), las variaciones de la asignación aleatoria pueden vencer los obstáculos políticos a las evaluaciones experimentales. Uno de estos diseños es el llamado diseño de “estimulación aleatoria”, el cual se aplica a programas voluntarios con tasas de participación menores al 100 por ciento. En este caso, en lugar de la aplicación aleatoria del tratamiento, se da un incentivo para participar tal como el subsidio o información adicional sobre el programa, se aplica de forma aleatoria a las unidades elegibles. En términos simples, esta estrategia crea un buen instrumento (ver la Sección 3.4 para mayor información sobre los instrumentos) a través de la creación de una variación aleatoria en la participación. Este método depende crucialmente de tener un incentivo que en efecto afecte la probabilidad de participar de manera mensurable. El segundo diseño alternativo consiste de un muestreo aleatorio en el margen, como en Black, Smith, Berger y Noel (2003). Ellos estudian el efecto de servicios obligatorios de recolocación para los receptores del seguro de desempleo en el estado de Kentucky. A los individuos se les asigna el servicio obligatorio con base en la duración predicha de su periodo de desempleo. Sólo los individuos al margen de ser tratados son seleccionados aleatoriamente. Esto prueba ser mucho menos profundo que los muestreos aleatorios de escala completa y también satisface la preocupación del estado para tratar a todos los solicitantes que han esperado durante mucho tiempo. En un mundo de efectos heterogéneos, ninguna de estas versiones alternativas de muestreo aleatorio estima el impacto medio del tratamiento en las unidades tratadas. Sin embargo, los parámetros que sí estiman pueden ser de gran interés para las políticas, si la preocupación política se centra en expansiones o contracciones marginales del programa.

En conjunto, los experimentos tienen un enorme poder, tanto por como sus propiedades estadísticas por sus propiedades retóricas. Los diseñadores de políticas, los expertos y cualquier persona pueden comprender los diseños experimentales, algo que no es muy común en los métodos no experimentales como el emparejamiento, variables instrumentales o modelos estructurales de equilibrio general. En contraste a esta situación, donde los evaluadores deben apremiar y espiar constantemente a las agencias renuentes para emprender

evaluaciones con asignación aleatoria, en una política debidamente ordenada, los funcionarios del gobierno tendrían la responsabilidad de formar una coraza para no aplicar el muestreo aleatorio en el caso de programas costosos o importantes que justifican una evaluación a escala completa y que no caen en las categorías inapropiadas descritas antes en esta sección.

3.2. Selección sobre observables: regresión y emparejamiento

La selección sobre observables ocurre cuando las características observadas determinan la participación en un programa pero, condicionando a aquellas características en las cuales, la participación no depende de los resultados ante la ausencia de la participación. En tales situaciones, condicionar sobre las características que determinan la participación basta para solucionar el problema del sesgo de selección. En general, la selección sobre observables tiene la mayor verosimilitud cuando los datos observados contienen variables que se relacionan con todos los factores principales identificados por la teoría (y la evidencia en programas similares) que afectaron tanto la participación como los resultados. Un ejemplo simple aclara este punto. Suponga que tanto hombres como mujeres eligen participar aleatoriamente en un programa de capacitación, pero los hombres eligen participar con una probabilidad más alta que la de las mujeres. Supongamos también que las mujeres tienen mejores resultados en el mercado laboral sin entrenamiento, que los hombres (una suposición no realista en las poblaciones localizadas por muchos programas de entrenamiento) y que el entrenamiento tiene el mismo efecto promedio en los hombres que las mujeres. La simple comparación de resultados de todos los no-participantes elegibles subestimará el impacto del programa, porque esta comparación combina el impacto de entrenamiento con el efecto de la sobre-representación de los hombres en el programa. Debido a que los hombres tienen peores resultados en el mercado laboral en ausencia de entrenamiento que las mujeres, la sobre-representación de los hombres en el programa combinará esta simple comparación de impacto en el programa bajo sesgo. Si en su lugar, separadamente se compara a los participantes y no participante masculinos con las participantes y no participantes femeninas, obtendremos una estimación insesgada del impacto del programa.

Con mucho, la manera más común de tomar en cuenta la selección en el tratamiento con base en características observables consiste en usar métodos de regresión lineal estándar, o sus análogos como los modelos logit y probit para variables dependientes limitadas, e incluyendo lo observado en el modelo. Una formulación estándar tendría el siguiente aspecto

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i$$

donde Y_i es el resultado de interés, D_i es una variable dicotómica por haber recibido el tratamiento (con β_D como el efecto correspondiente al efecto de tratamiento),

X_{i1}, \dots, X_{ki} son las variables confusas en los que la regresión será estimada con una muestra de unidades tratadas y elegibles no tratadas (lo que significa que no puede usarse este enfoque para un tratamiento que se llega a todas las unidades elegibles). En un mundo de “efectos comunes” se mantienen los supuestos de la selección por observables, β_D estima el efecto común de tratamiento. En un mundo con efectos heterogéneos de tratamiento, se estima el impacto de tratamiento en las unidades tratadas bajo suposiciones generales.

La regresión tiene grandes ventajas por su familiaridad y facilidad para la interpretación y uso. Todos los paquetes estadísticos estándares la incluyen, e incluso, algunos programas de bases de datos. Los coeficientes tienen interpretaciones como derivadas parciales o diferencias finitas (a través de ellos surgen los modelos un poco más complicados logit y probit, algunos paquetes estadísticos ahora reportan los efectos marginales, los cuales son primos cercanos de las derivadas parciales.)² A pesar de esas ventajas es importante notar que “la regresión en general no es una técnica con un impacto neto expedito y satisfactorio”, como declara el capítulo de Eberts y O’Leary en este volumen. Si la regresión produce o no estimaciones consistentes depende de si los supuestos de la selección de observables se mantuvieron, de lo que dependerá por tanto, la riqueza del conjunto de variables disponibles para incluir en la regresión y la naturaleza del proceso de selección en cada contexto. Como se enfatiza en la sección 3.7 de abajo, un estimador no econométrico brinda una solución general al problema de la selección de la muestra. En algunos casos la regresión hará eso, en otros no. Una de las principales contribuciones del evaluador consiste en determinar que caso corresponde a la evaluación que tiene a la mano, una tarea que requiere más que frases calmantes.

Adicionalmente a los supuestos de la selección en observables, el análisis de regresión lineal estándar impone también una forma funcional lineal en los datos, la cual puede o no corresponder a las relaciones de una población subyacente. Incluyendo un mayor orden en términos de X_i relaja sus restricciones, pero esto raramente se hace en la práctica. Los métodos de emparejamiento, los cuales han recibido mucha atención en la bibliografía sobre evaluación en los últimos años, relajan los supuestos de la forma funcional lineal, inherentes en el enfoque de regresión estándar, al mismo tiempo mantienen el supuesto de selección en observables.

La idea básica del emparejamiento consiste en comparar directamente a los individuos con exactamente los mismos (o similares) valores para las variables confusas relevantes. Esto evita cualquier forma funcional de restricciones. El modo más fácil de ver esto es considerar un ejemplo con solo variables discretas X . En ese caso, la versión más simple de estimador de emparejamiento consiste en encontrar, para cada unidad tratada, una unidad no tratada con valores idénticos

en sus covariables. El impacto estimado, el cual provee un estimado del impacto del tratamiento sobre los tratados en un mundo de efectos heterogéneos, consiste entonces el resultado medio para las unidades tratadas menos el resultado medio para las unidades emparejadas no tratadas. Un indicador de la ventaja del efecto de tratamiento puede obtenerse al revalorar los estimadores específicos de X a través de una distribución de la población X (en lugar de ponderarlos implícitamente a través de la distribución de X sobre el grupo de tratamiento, como normalmente se realiza en el emparejamiento).

Cuando el conjunto de las variables de emparejamiento incluyen variables continuas, o un gran número de variables discretas, el emparejamiento exacto se dificulta.³ En dichos casos, el emparejamiento depende de la medición de la distancia para determinar cuales de las observaciones no tratadas podrían jugar un papel en la estimación del contrafactual por cada observación tratada. Puesto de manera diferente, la distancia métrica convierte distancias de un vector de variables en un solo número, el cual puede ser usado para emparejar observaciones similares. En el caso de una sola variable continua, las diferencias absolutas en esa variable pueden utilizarse como parámetros. Con un conjunto de covariados más riguroso, las opciones incluyen emparejar el sistema métrico de Mahalanobis, como en Rosenbaum y Rubin (1985), y resultados propensos de emparejar, como se desarrolla en Rosenbaum y Rubin (1983). Los resultados propensos son sólo la jerga estadística para la probabilidad de participación condicional en X .

La bibliografía económica tiende a enfocarse en el resultado propenso de emparejamiento debido a que la teoría proporciona con frecuencia una guía sobre qué variables deberían afectar la probabilidad del tratamiento y por qué la probabilidad de participación es con frecuencia de interés independiente, de modo que sería calculado de todas formas. Con los resultados propensos de emparejamiento, el contrafactual estimado para una supuesta observación tratada está basado en los resultados de las observaciones no tratadas con una probabilidad de participación similar. Con el propósito de satisfacer los supuestos de la selección sobre observables, el modelo de probabilidad de participación debería incluir todas las variables importantes que afectan tanto la participación como el resultado —no solamente aquellas en el conjunto de datos establecidos y a la mano.

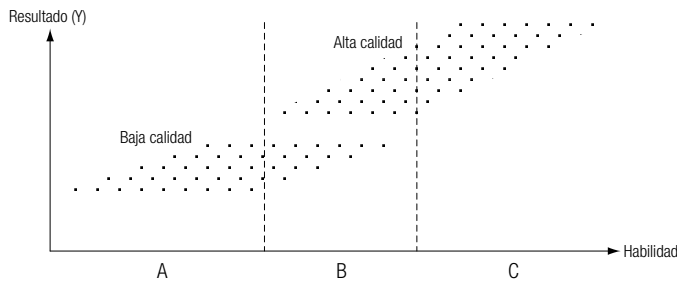
Los métodos de emparejamiento tienen el beneficio adicional, con relación a los métodos de regresión estándar, de enfocar la atención a los llamados problemas de “soporte”. Los problemas de soporte se presentan cuando los datos no cuentan con observaciones similares no tratadas para algunas de las observaciones tratadas. En tales situaciones, es usual que al aplicar los métodos de emparejamiento, simplemente para eliminar del análisis dichas observaciones tratadas, lo cual puede afectar sustancialmente la naturaleza del parámetro que está siendo estimado, si hay muchas de ellas. En contraste, los métodos de

regresión estándar producirán estimaciones incluso en la ausencia de unidades de comparación que parezcan las unidades de tratamiento, debido a que la forma funcional lineal llena la información faltante. Dicho de otra forma, la regresión identifica el resultado del modelo de resultados sin tratamiento en la región de los datos donde las unidades no tratadas están situadas, y entonces las proyecta hacia la región de los datos en donde las unidades tratadas están situadas, y de esta forma estiman implícitamente el contrafactual.

La Figura 12.1 ayuda a ilustrar este punto. El eje horizontal representa la variable de emparejamiento bajo el supuesto que condicionar en esta única X toma en cuenta la selección no aleatoria en el tratamiento. El eje vertical representa los resultados. Las dos nubes de datos representan a las unidades tratadas y no tratadas respectivamente. La Región B, en el centro de la figura, constituye la región de “soporte común”, donde existen tanto observaciones tratadas como no tratadas con valores de X más o menos similares (lo suficientemente cercanos para ser “buenos” emparejamientos). La condición de soporte recae en la Región C, en el lado derecho de la figura, que incluye observaciones tratadas pero no las no tratadas con valores similares a los de X . No puede construirse una estimación del objetivo por emparejamiento para las observaciones tratadas en la Región C. Por el contrario, no existiría problema de incluirlas en un análisis de regresión, cuya función lineal proyectaría el resultado condicional del medio sin tratamiento que se estima usando los datos en las unidades no tratadas en las Regiones A y B hacia la Región C. A fin de cuentas, las secciones de este tipo pueden funcionar en un contexto dado, pero el evaluador todavía desearía saber si sus estimaciones confían intensamente en la función lineal o no. Finalmente, la Región A contiene unidades no tratadas pero no unidades tratadas. Cuando se estima el impacto del tratamiento en las tratadas, esto no plantea problema alguno; el emparejamiento ignora estas unidades, ya que no se requieren para construir el contrafactual para cualquiera de las unidades tratadas. Por otro lado, los métodos de regresión lineal estándar hacen uso de estas observaciones para fijar la relación entre X y el resultado. Ver Heckman y Vytlacil (2001a), y Black y Smith (2004) para una explicación detallada del problema de soporte. Ya sea que el emparejamiento se realice con base en $P(X)$ o alguna otra distancia métrica basada en X , existe una variedad de métodos para realizar el emparejamiento cuando los datos carecen de parejas exactas. Estos métodos difieren en varias dimensiones, de las cuales quizás las dos más importantes son si ellos utilizan o no cada observación no tratada para apoyar la construcción de la estimación del contrafactual para más de una observación tratada. La forma más simple de emparejar —común en la bibliografía de estadística aplicada pero no en la bibliografía de economía— es emparejar a un vecino cercano sin reemplazo. En este caso, una sola observación no tratada estima el contrafactual para cada observación tratada, y las observaciones no tratadas sólo pueden estimar el contrafactual para una sola observación tratada, aunque sea la observación no tratada más cercana para

muchas observaciones tratadas. Este método tiene la desventaja de que tiende a desechar una gran cantidad de observaciones de comparación y de que, en conjuntos de datos con sólo unas pocas observaciones de comparación, tiende a conducir hacia malos emparejamientos, en el sentido de emparejar observaciones tratadas con observaciones no tratadas que no se parecen entre sí. Dehejia y Wahba (1999) proporcionan una exposición útil de este asunto, con un ejemplo empírico que ilustra los peligros de emparejar sin reemplazar.

Figura 12.1



Otras versiones de emparejamiento incluyen el núcleo del emparejamiento, el emparejar con el vecino más cercano y con más de un vecino cercano (esto incrementa el sesgo pero disminuye la varianza), el emparejamiento local lineal y la ponderación por el inverso de la propensión del resultado. Ver Smith y Todd (2004) para una revisión relativamente aplicada de los diferentes métodos y, entre otros, a Heckman, Ichimura y Todd (1997, 1998), Heckman, Ichimura, Smith y Todd (1998), y Hahn (1998) para los detalles técnicos. Imbens (2004) proporciona una encuesta legible de la bibliografía técnica reciente mientras que Heckman y Navarro-Lozano (2004) contrastan métodos de emparejamiento por probabilidad de participación con métodos basados en restricciones de exclusión y explican la probabilidad estimada del tratamiento en el papel diferencial que desempeña cada una. Frölich (2004) presenta un análisis de Monte Carlo que sugiere que el núcleo del emparejamiento (y otro método denominado cotejo de sistemas) tiende a superar en ejecución al emparejar con el vecino más cercano, el emparejamiento local lineal y la ponderación por el inverso del estimado de la propensión del resultado. Lechner (1999, 2000) proporciona algunos buenos ejemplos del emparejamiento en la práctica.

En la bibliografía existe una variedad de modalidades para el emparejamiento. Por ejemplo, los datos longitudinales permiten la combinación de emparejar con los métodos de la diferencia-en las-diferencias expuestos en la Sección 3.4.

Heckman, Ichimura, Smith y Todd (1998) desarrollan estos métodos. Ver Blundell y Costa Dias (2002) para una introducción menos técnica (que también incluye el caso de datos cruzados repetidos) y a Eichler y Lechner (2002) para una aplicación. En estudios de programas que puedan crear efectos a nivel de un pueblo o una ciudad pequeña, se toman en cuenta los diseños de emparejamiento en áreas locales. Estos diseños generalmente comparten el problema de tener muy pocas áreas tratadas y no tratadas para efectuar inferencia estadística confiable. Ver Long y Wissoker (1995) como un ejemplo.

En resumen, emparejar representa una importante ampliación de los enfoques de regresión lineal tradicionales cuando los datos se apoyan en los supuestos de la selección en observables. Emparejar no soluciona el problema de selección sobre no observables, pero proporciona flexibilidad para condicionar y facilita examinar el aspecto de respaldo. El emparejar, ya sea semi-paramétrico (probabilidad de participación) o no paramétrico (emparejamiento de células), requiere de más datos que los enfoques tradicionales que imponen más estructura al problema. Aunque los paquetes estadísticos estándar todavía no contienen rutinas para efectuar el emparejamiento, existen varias rutinas escritas para los usuarios del Stata.

3.3. Selección sobre no observables: métodos longitudinales

Un método muy simple para evaluar una política de desarrollo consiste en examinar las diferencias en los resultados de las unidades afectadas por la política antes y después de que se aplique. El supuesto subyacente implícito en esta estrategia simple es que las unidades sujetas a un cambio de política habrían tenido el mismo resultado que antes, si la política no hubiera tenido una intervención para ser cambiada. Aunque razonable, en algunos contextos esta suposición requiere que las unidades tratadas no elijan su entrada al tratamiento con base en cambios temporales en sus resultados. Por ejemplo, si las empresas sólo eligen participar en un programa de subsidios cuando están teniendo un mal año, y si la mayoría de los años malos son seguidos por años de bonanza sin el subsidio, entonces una comparación antes y después de los resultados de las compañías participantes exagerará el impacto del subsidio en el desempeño de la empresa al atribuirle el rendimiento normal de los tiempos de bonanza de la empresa subsidiada. El estimador antes-después también requiere de la ausencia de cambios agregados en los resultados debido a la macroeconomía u otros factores. Si la economía funciona bien y todas las unidades mejoran su desempeño, entonces el estimador asignará incorrectamente las ganancias al tratamiento.

En términos de anotación, dejemos que t indique un periodo posprograma y t' un periodo previo al programa. Entonces el estimador antes-después se da:

$$\Delta_{BA} = E(Y_{1t} | D=1) = E(Y_{1t'} - Y_{0t'} | D=1).$$

El estimador es consistente si y sólo si $E(Y_{0t} | D=1) = E(Y_{0t} - Y_{0t} | D=1)$, que es tan sólo la condición de estabilidad en el párrafo anterior expresada en la anotación. Observe que la selección no aleatoria de unidades del tratamiento no representa problema alguno mientras que la condición de estabilidad sea satisfactoria, porque el estimador antes-después compara las unidades tratadas con ellas mismas antes del tratamiento. El estimador antes-después estima el impacto del tratamiento en las unidades tratadas y no captura efecto alguno del equilibrio general, a menos que sea incluido en los resultados usados para calcular las estimaciones (como por ejemplo variables de resultados a nivel ciudad). Los diseños de series de tiempo interrumpidos, denominados en la bibliografía empírica financiera como análisis de eventos históricos, generalizan el estimador antes-después al incluir periodos adicionales de datos antes o después del tratamiento de interés. Esto permite al investigador el tener un mayor control para el caso de tendencias preexistentes en los resultados.

La preocupación acerca de confundir los efectos del tratamiento con los cambios generales en la economía motivan al estimador denominado “diferencia en las diferencias”. Este estimador compara el cambio antes y después de las unidades tratadas con el cambio antes y después en las unidades no tratadas. Al hacerlo así, cualquier tendencia común, que aparezca en los resultados tanto de las unidades no tratadas como las tratadas, será independiente. En términos de nuestra anotación, el estimador de la diferencia en las diferencias consiste en:

$$\Delta_{DD} = [E(Y_{1t} | D=1) = E(Y_{0t} | D=1)] - [E(Y_{0t} | D=0) = E(Y_{0t} | D=0)].$$

La tendencia de tiempo común justifica el estimador proporcionado por:

$$E(Y_{0t} | D=1) = E(Y_{0t} | D=1) - E(Y_{0t} | D=0) = E(Y_{0t} | D=0)]$$

Comunmente los investigadores estimarán la diferencia en las diferencias utilizando un modelo de regresión relativamente simple, como en:

$$Y_i = \beta_0 + \beta_T T_i + \beta_D D_i + \beta_{DD} T_i D_i + \varepsilon_i,$$

donde $T_i = 1$ en el periodo “t” y $T_i = 0$ en cualquier otro caso y donde omitimos la variable X (de la notación, no del modelo) por simplicidad. El coeficiente β_T mide el efecto de la tendencia común en el tiempo, mientras que β_D estima la diferencia constante del tiempo en los resultados no tratados entre las unidades tratadas y no tratadas, y β_{DD} proporciona la estimación del impacto de la diferencia en las diferencias. En un mundo de efectos comunes (el cual es un mundo de efectos homogéneos del tratamiento), β_{DD} estima el impacto medio del tratamiento en las unidades tratadas (y todos los otros parámetros de impacto medio). En un mundo de efectos heterogéneos, β_{DD} estima el impacto del significado del tratamiento en las unidades tratadas bajo suposiciones bastante generales. Así como el estimador antes-después, el estimador de la diferencia en las diferencias generaliza para el

caso de más periodos de tiempo ya sea antes o después del tratamiento o ambos casos. Blundell, por ejemplo (2001) expone e implementa dichos estimadores.

Los modelos de panel de datos constituyen la versión más general de estos estimadores. Estos modelos se aplican a conjuntos de datos con múltiples observaciones a lo largo del tiempo sobre muchas unidades tratadas (y quizás algunas unidades no tratadas). Se corre una regresión del resultado-variable de interés sobre los covariados exógenos además de variables dicotómicas para cada unidad y cada periodo de tiempo. Las unidades variables dicotómicas controlan las diferencias permanentes en los resultados entre las unidades, como en el modelo simple de la diferencia en las diferencias. Las dicotómicas en un periodo controlan los efectos agregados en cada periodo. Los modelos de panel requieren algunas variaciones en el momento del tratamiento, sin dicha variación, los efectos del tratamiento no pueden ser distinguidos de los efectos agregados en el tiempo.⁴ Estos modelos también requieren que el tiempo del tratamiento de las unidades no dependa de cambios transitorios en los resultados. Esta no es una suposición inocua. Como Heckman y Smith (1999) demuestran en el contexto de un programa gubernamental de capacitación para el trabajo, la participación individual depende de manera crítica de los *shocks* temporales del mercado laboral, con el resultado que los estimadores longitudinales tienen un mayor sesgo.

En los términos de la anotación antes presentada, el modelo básico de panel presenta la siguiente forma:

$$Y_{it} = \beta_0 + \beta_D D_{it} + \mu_i + \mu_t + \epsilon_{it},$$

donde β_D es el estimador de impacto del panel de datos, D_{it} es un indicador que varía en el tiempo para el tratamiento, μ_i es la intercepción de una unidad específica, μ_t es la intercepción sobre un periodo de tiempo específico y nuevamente omití la X por simplicidad. La intercepción sobre un período de tiempo absorbe cualquier tendencia común, mientras que la intercepción de cada unidad específica absorbe las diferencias no variables entre las unidades. Esencialmente, lo que queda es un promedio ponderado de las estimaciones antes-después para las diferentes unidades tratadas.

Un par de ejemplos ilustra cómo funcionan los modelos. Evans y Topoleski (2002) evalúan el impacto de los casinos de los nativos americanos respecto al empleo y otros resultados (tales como la criminalidad y las quiebras de bancos en el mismo país) usando métodos de datos de panel. Ellos construyen los datos de panel sobre los resultados de las tribus de nativos americanos en los Estados Unidos que tienen y sobre los que no tienen casinos. Combinan esto con datos sobre el momento de la apertura del casino para aquellas tribus que poseen uno. Debido a la complicada estructura legal que rodea al juego en los casinos de los Estados Unidos, y dado el hecho de que tanto el gobierno estatal como el

federal desempeñan un papel, existe una gran variación tanto en la incidencia como el momento de apertura de los casinos entre las tribus. Es más, la variación en el momento de apertura de los casinos entre las tribus que los poseen está plausiblemente poco relacionada con el esquema temporal de resultados en la ausencia de los casinos. De una manera más simple, el momento de apertura depende de las vacilaciones de los sistemas político y legal, y no de los cambios en la tasa de empleo tribal. La variable dicotómica tribal incluida en este modelo toma en cuenta las diferencias permanentes entre las tribus, y las variables dicotómicas por año, toman en cuenta las tendencias nacionales que afectan a todas las tribus. Ellos encontraron que los casinos incrementan el empleo tribal en la misma tasa que en la población, así como que incrementan el empleo en los alrededores del condado (así como el crimen y la quiebra de bancos).

Coates y Humphreys (1999) proporcionan otra aplicación al desarrollo económico de los métodos de datos de panel. Ellos recopilan datos acerca de las ciudades en los Estados Unidos, incluyendo la presencia o ausencia anual de franquicias de fútbol, béisbol y básquetbol profesional y de la construcción de nuevos estadios y arenas. Usando el mismo marco básico de trabajo descrito anteriormente, ellos examinan el efecto de las franquicias en favor de los deportes y de la construcción de nuevas instalaciones deportivas (casi siempre con un monto sustancial de dinero público y casi siempre justificados como motores del crecimiento económico) tanto en los niveles como en las tasa de crecimiento de la actividad económica. Ellos encontraron poco sobre los efectos económicos del gasto público en beneficio de los equipos deportivos, lo cual se comporta conforme a la bibliografía seria del resto del tema; ver por ejemplo los artículos recolectados en Noll y Zimbalist (1997).

En general, los métodos de datos de panel representan una poderosa herramienta cuando están disponibles los datos longitudinales sobre las unidades tratadas y no tratadas, y cuando el momento de la aplicación del tratamiento varía entre las unidades, al igual que cuando el momento de la aplicación de tratamiento no está relacionado con los resultados que están siendo estudiados, condicionados por las variables condicionales incluidas. Esta última condición es algunas veces ignorada en la bibliografía. Los datos adicionales permiten la comprobación de los supuestos en muchos contextos, véase por ejemplo Moffitt (1991) y Heckman y Hotz (1989). Para los lectores interesados en los detalles potenciales de los métodos de datos de panel, la muy controversial bibliografía sobre los impactos de las leyes estatales del “derecho a portar armas” (éstas permiten que determinados tipos de ciudadanos estadounidenses puedan portar armas ocultas), la cual se apoya casi por completo en tales métodos, profundiza en qué puede salir bien o qué puede salir mal con este enfoque. Ver por ejemplo, Lott y Mustard (1997), Black y Nagin (1998), y Ayres y Donahue (2002).

3.4. Selección sobre no observables: variables instrumentales

Los métodos basados en variables instrumentales (VI) o restricciones de exclusión representan una estrategia econométrica alternativa para enfrentarse con la selección sobre no observables. Un instrumento, o restricción de exclusión, es una variable que afecta la participación en el tratamiento pero que no afecta el resultado más que su efectividad en la participación en el tratamiento (condicionada en las otras variables incluidas en la ecuación del efecto-resultado). El uso de “restricción de exclusión” surge del hecho que tales variables pueden ser excluidas de la ecuación del efecto-resultado, pero incluidas en la ecuación del tratamiento. A diferencia de los métodos longitudinales, los métodos VI requieren sólo datos de corte transversal y potencialmente pueden enfrentar la selección de no observables que varíe con el tiempo. Véase Angrist y Krueger (2001) para una exposición más amplia pero no técnica de los métodos VI, y véase a Heckman, LaLonde y Smith (1999) para tratamientos más técnicos.

Un ejemplo simple con un instrumento binario en un mundo de efectos comunes proporciona la idea básica. Considere dos pueblos idénticos salvo que uno está ubicado cerca de un centro público de capacitación y el otro está alejado del mismo centro de capacitación. Cada pueblo incluye 100 personas elegibles, de las cuales 50 tienen auto. El resultado en ausencia de la capacitación es igual a 100 para todas las personas elegibles. El beneficio de la capacitación (neto del costo de oportunidad por el tiempo del participante) es igual a 10 para todos (un mundo de efectos comunes), mientras que el costo de aquellas en el pueblo cercano para llegar al centro de capacitación es igual a 1. Para las personas elegibles en el pueblo lejano, el costo de transporte al centro de capacitación es igual a 5 para aquellas con auto y de 15 para aquellas sin auto. La realización de todo lo anterior resulta en que todas las personas elegibles en el pueblo cercano toman la capacitación pero que únicamente las elegibles con auto en el pueblo lejano se capacitan. En consecuencia, podemos utilizar la ubicación como un instrumento, porque ésta afecta la participación en el tratamiento —una fracción diferente de los participantes elegibles en cada pueblo— pero no afecta los resultados aparte de los efectos del tratamiento —todos obtienen 100 en ausencia del tratamiento en ambos pueblos—.

El estimador Wald para instrumentos binarios, es:

$$\Delta_{VI} = \frac{E(Y | Z = 1) - E(Y | Z = 0)}{\Pr(D = 1 | Z = 1) - \Pr(D = 1 | Z = 0)}$$

es suficiente para nuestro sencillo ejemplo (aunque los mínimos cuadrados en dos etapas podrían también ser utilizados y producirían una estimación equivalente). En el estimador Wald, como antes, Z denota el instrumento y D denota la participación en el tratamiento; en nuestro sencillo ejemplo, $Z=1$ para

las unidades elegibles en el pueblo cercano y $Z=0$ para los individuos en el pueblo lejano. Incorporando las cifras del ejemplo produce

$$\Delta_{VI} = \frac{[100 + (1.0)(10)] - [100 + (0.5)(10)]}{1.0 - 0.5} = \frac{5}{0.5} = 10,$$

la cual es la respuesta correcta. Este ejemplo ilustra el punto clave de que una variación en la recepción del tratamiento que no se relaciona con los resultados en ausencia del tratamiento, el instrumento identifica el efecto del tratamiento a través de la comparación de los resultados de grupos con diferentes valores del instrumento.

Cuando el instrumento es continuo o variado, en lugar de binario, o cuándo múltiples instrumentos están disponibles (lo que en verdad representa buena fortuna), el estimador Wald es reemplazado por el método estándar de mínimos cuadrados en dos etapas. En el estimador en dos etapas, la variable “endógena” (participación en el tratamiento) primero experimenta retroceso acerca del instrumento o instrumentos y cualquier variable X exógena en el modelo. En la segunda etapa, el resultado está sujeto al retroceso sobre los valores predichos de la primera etapa, al igual que la variable X . Intuitivamente, utilizar las predicciones de la primera etapa, en lugar de la variable dicotómica de participación de sí misma, omite la variación en la participación que no proviene de los factores sin relación con los resultados en ausencia del tratamiento. Muy poca bibliografía en econometría explora estimadores de evaluación semi-paramétricos menos restringidos basados en restricciones de exclusión, aunque estos estimadores han tenido muy poco uso en el trabajo aplicado. Ver Newey, Power y Walter (1990), y Blundell y Powell (2001) para mayor explicación.

Los instrumentos son una maravillosa herramienta cuando están disponibles, pero ¿de dónde provienen? La creación deliberada representa la manera más directa de obtener instrumentos. Una manera de pensar sobre los experimentos sociales es que son mecanismos para crear buenos instrumentos; esto incluye el diseño de estimulación aleatoria descrito en la Sección 3.1. Ver Heckman (1996) para más información sobre esta perspectiva de los experimentos. Una segunda forma de creación deliberada de instrumentos consiste en la combinación de teoría con la recolección inteligente de datos. Por ejemplo, Card (1995) agregó información sobre la distancia a la universidad o colegio más cercano a un conjunto de datos estándar a nivel individual y entonces utiliza las distancias como un instrumento en los años de escolaridad. La manera definitiva, y en la práctica, la más importante, de recolectar datos combina la teoría con el conocimiento institucional. Los cambios institucionales que aparentan carecer de relación con los resultados de interés (o al menos cuya ocurrencia no está relacionada con ellos, como por ejemplo, algunas decisiones de la corte) pero que afectan la participación en un tratamiento pueden proporcionar buenos instrumentos. La huelga de autobuses utilizada para

aportar variación en el cuidado prenatal, en Evans y Lien (2002) proporciona tal ejemplo. Las diferencias en las decisiones de los gerentes del programa o en la intensidad del programa entre las jurisdicciones constituyen una fuente potencial de instrumentos en muchos contextos de desarrollo económico. En general, los evaluadores deberían considerar el hallazgo de los instrumentos como un beneficio colateral al reflexionar sobre la economía de un problema de evaluación dado y de aprender sobre las instituciones relevantes, ambas actividades necesarias por su propio derecho.

El modelo de selección normal bivariada de Heckman (1979) está relacionado estrechamente con el modelo de variables instrumentales. El modelo normal bivariado, como su nombre lo sugiere, asume que los términos de error en la ecuación de resultados y de participación tienen una distribución normal bivariada. Bajo esta suposición, Heckman (1979) proporciona un estimador en dos etapas que estima el impacto del tratamiento sobre las unidades tratadas en un mundo de efectos comunes.⁵ Formalmente, el modelo normal bivariado no requiere de una restricción de exclusión; las suposiciones de la forma funcional son suficientes para identificar el parámetro de interés. Sin embargo, es bien conocido en la bibliografía (ver, por ejemplo, la encuesta de Puhani, 2000) que el modelo tiende a la inestabilidad sin una restricción de exclusión. Por lo tanto, los evaluadores deberían evitar este estimador a menos que tengan a la mano un buen instrumento. De manera más general, como observan Heckman y Robb (1985), este estimador requiere suposiciones más fuertes que el estimador del VI; si el supuesto de los efectos comunes es posible en un contexto dado, entonces se prefiere el estimador de VI.

Las cosas se vuelven más complicadas en un mundo con efectos de tratamiento heterogéneos, particularmente cuando esos efectos de tratamiento están correlacionados con el instrumento. Para ver esto, regrese al ejemplo simple de los dos pueblos que comparten un centro de capacitación considerado anteriormente. Suponga que la mitad de las personas elegibles para el programa en cada pueblo tienen un beneficio de 10 del programa, mientras que la mitad tiene un beneficio de cinco. Ahora, suponga también, que los costos de transporte son homogéneos, de modo que todos en el pueblo cercano tienen un costo de cero para llegar al centro de capacitación, mientras que todos en el pueblo lejano tienen un costo de siete. En esta versión de la historia, todos en el pueblo cercano participan de la capacitación nuevamente, junto con la mitad de las personas elegibles en el pueblo lejano. Sin embargo, ahora en lugar de que las personas elegibles del pueblo lejano con costos de transporte bajo participen, son los elegibles del pueblo lejano con impacto de 10 los participantes. Para ellos, el impacto de 10 excede el costo de transporte de siete, mientras que para las personas elegibles con un impacto de cinco en el pueblo lejano, el costo de transporte de siete hace que no valga la pena su participación.

¿Cómo es que estos cambios en la historia de los dos pueblos afectan el estimado producido por el estimador Wald? El impacto significativo en el pueblo cercano es ahora $107.5 = [100 + (0.5 \cdot 10) + (0.5 \cdot 5)]$. El resultado significativo en el pueblo lejano es igual a $105 = [100 + (0.5 \cdot 10)]$. La probabilidad de participación en el pueblo cercano es igual a 1.0 y la misma en el pueblo lejano es igual a 0.5. Así, la estimación de Wald es igual a $(107.5 - 105) / (0.5) = 2.5 / 0.5 = 5.0$. Esta estimación podría sorprender, dado que la mayoría de quienes participan (dos tercios, para ser más precisos) reciben un impacto de 10.

La característica clave de esta versión de la historia yace en la correlación entre los impactos (condicional en la participación) y el instrumento. En el pueblo cercano el impacto significativo entre los participantes es 7.5, mientras que en el pueblo lejano es de 10. Cuando un instrumento binario, tal como éste, está correlacionado con los efectos heterogéneos del tratamiento en forma condicional sobre la participación, el estimador Wald ya no estima más el impacto significativo del tratamiento en las unidades tratadas. En su lugar, se estima lo que Imbens y Angrist (1994) denominan un Efecto Local Promedio del Tratamiento (LATE, por sus siglas en inglés). En palabras llanas, el estimador Wald ahora estima el impacto significativo sobre aquellas unidades que cambian su estado de participación en respuesta al cambio en el valor del instrumento. Las unidades que cambiaron su estado de participación cuando el instrumento cambia en la historia de los dos pueblos son aquellas con impacto de cinco; participan en el pueblo cercano pero no en el lejano. Así, el LATE en este caso se equipara a cinco.

Observe que los otros parámetros estándar del tratamiento no son iguales a cinco en este caso. Por ejemplo, el impacto significativo del tratamiento de las unidades tratadas es igual $(2/3) \cdot 10 + (1/3) \cdot 5 = 25/3 = 8.33$. El efecto promedio del tratamiento es igual a 7.5, debido a la suposición de que la mitad de las unidades elegibles en cada pueblo tienen un impacto de 10 y la otra mitad un impacto de cinco. La relación entre el tratamiento en las unidades tratadas y el LATE es instructiva aquí. El tratamiento en los parámetros tratados excede el LATE debido a los participantes inframarginales —aquellos que participan sin importar el valor del impacto (en este caso, aquellos con impacto de 10) tienen impactos significativos más altos que los participantes marginales, como lo predice la teoría económica. Los tres parámetros difieren en este caso y responden diferentes preguntas de interés sobre políticas. Heckman (1997), Angrist y Krueger (1999) y Heckman, LaLonde y Smith (1999) exponen con mayor detalle el caso de los instrumentos binarios.

El artículo de Sweetman, Warburton, McPhee y Warburton (2003) proporciona una buena aplicación del ejemplo de la estimación del LATE. En su estudio, el instrumento consiste en la persona que toma la decisión final en los casos de beneficios para discapacitados en la provincia canadiense de la Columbia Británica. Cuando la persona que toma las decisiones cambia y la probabilidad

de aceptación incremental, los autores pueden estimar un LATE muy interesante, es decir el impacto recibe pagos por discapacidad en los candidatos marginales cuya aplicación habría sido rechazada bajo un régimen, pero aceptados bajo el otro. Por suposición los mismos autores no pueden estimar el impacto de recibir pagos del seguro por discapacidad de los solicitantes inframarginales cuyos casos serían aprobados bajo ambos regímenes.

Las cosas se vuelven un poco más complicadas con instrumentos continuos, o con instrumentos múltiples, o con tratamientos continuos (o multivariados) en un mundo de efectos de tratamiento heterogéneo. Ver Angrist e Imbens (1995), y Heckman y Vytlačil (1998) para una exposición de los tratamientos multivariados. El modelo de selección normal bivariada también se generaliza para el caso del tratamiento heterogéneo y ha recibido mucha atención en la bibliografía. En este caso, la estructura adicional proporcionada por el fuerte supuesto de distribución en los términos de error, permite la estimación de numerosos LATEs, así como el ATE y el tratamiento en el parámetro tratado. Ver Heckman, Tobias y Vytlačil (2003) para mayores detalles y un ejemplo empírico. Recientemente ha aparecido en la bibliografía una versión semiparamétrica del modelo de selección normal de bivariada. Una descripción completa de este modelo está más allá del nivel de este artículo de investigación, ver Hackman y Vytlačil (2001b) para más detalles.

En resumen, los instrumentos representan una poderosa herramienta para derivar interesantes estimaciones de los impactos de los programas de desarrollo económico local cuando están disponibles. A la fecha, los métodos por VI han sido utilizados muy poco en la bibliografía de desarrollo económico; por lo mismo, los investigadores avisados, que están en búsqueda de novedosas fuentes de variación exógena, probablemente tienen al alcance de la mano un fruto empírico que cosechar para aumentar nuestro conocimiento tanto de los impactos de las políticas de desarrollo económico como del conocimiento del comportamiento de estos estimadores en la práctica.

3.5. Diseños discontinuos

La discontinuidad de diseños aplica a tratamientos asignados utilizando una variable en particular o un conjunto de variables, en el cual se distinguen las unidades tratadas y no tratadas por una abrupta interrupción en el valor de las variables. De esta forma podemos imaginar un programa que proporciona educación remedial a los estudiantes de edad preescolar con base en su puntaje en una prueba estandarizada, donde aquellos por debajo de cierto puntaje reciben el tratamiento y aquellos más altos no. O las subvenciones para pueblos pequeños que pudieran estar disponibles únicamente para esos pueblos con poblaciones menores a 15000 habitantes. En general, imagínese que el tratamiento es

proporcionado a aquellas unidades con un valor de una variable Z mayor que un valor con límite C.

En ausencia de suposiciones adicionales, la discontinuidad de los diseños estiman el efecto promedio del tratamiento en las unidades localizadas en el límite C a través de la comparación de los resultados de las unidades justo arriba del límite C con los resultados de las unidades justo por debajo del límite. En términos de nuestra anotación, la discontinuidad de diseños estima

$$\Delta_{\text{DIS}} = E(Y_1 - Y_0 | X \gg C)$$

Por ejemplo, en el caso de las subvenciones para los pueblos pequeños, la discontinuidad del estimador estima la efectividad de las subvenciones en pueblos con poblaciones de 15 000 habitantes, mediante la comparación de los resultados de los pueblos por debajo de los 15 000 habitantes con los resultados de los pueblos justo por arriba de los 15 000 habitantes. La forma exacta de la comparación depende en particular de la discontinuidad del estimador seleccionado.

Por supuesto, que en un mundo de efectos comunes, el impacto estimado en la discontinuidad del diseño se generaliza a otras unidades. Al agregar el supuesto que los resultados de las unidades no tratadas son lineales en los covariados los lleva al denominado diseño “discontinuidad de la regresión”, el cual puede ser estimado utilizando los métodos de regresión estándar. Dependiendo de la aplicación, la discontinuidad de diseños puede capturar efectos de equilibrio general. Que lo hagan o no depende del grado en la cual los resultados en las unidades tratadas influyan en los resultados de las unidades no tratadas. Por ejemplo, en el caso de las subvenciones a los pueblos pequeños, el estimador recogerá cualquier efecto de equilibrio general reflejado por las variables de resultados a nivel de los pueblos, pero será sesgado por la migración de los pueblos grandes hacia los pequeños en respuesta a las subvenciones.

La clave de la discontinuidad del diseño es que las unidades (u otros) no deberían poder manipular Z con el propósito de hacer que ciertas unidades sean tratadas y otras no. Algunos tratamientos cumplirán con este requerimiento y otros no. El ejemplo antes proporcionado acerca de una regla de tratamiento basada en una prueba estandarizada, sin componentes subjetivos, es un ejemplo de lo primero. Una política que proporciona beneficios a empresas con menos de cinco empleados proporciona un ejemplo de lo segundo. Un beneficio suficientemente atractivo hará que las empresas con seis o más empleados recorten su nómina a cinco. Es casi seguro que estas empresas serán un ejemplo aleatorio de las empresas con seis o más empleados. Como resultado, la comparación al margen entre las empresas con cinco y seis empleados ya no estimarán el parámetro de interés.

La bibliografía proporciona algunas presentaciones teóricas que podrían ser útiles. Ver por ejemplo, la sección con relación al tema en Heckman, LaLonde y Smith (1999), así como, Hahn, Todd y van der Klaauw (2001), a un nivel más técnico. El artículo posterior enfatiza la diferencia entre sesgo y discrepancia relacionados con la decisión de cuánto peso asignar a las observaciones a cierta distancia desde el punto de corte. Las solicitudes incluyen a Van der Klaauw (2002), quien estudia el efecto de la ayuda financiera en las admisiones a las universidades, Black (1999), estudia los precios habitacionales en los límites de las áreas de asistencia escolar (entre los distritos escolares), y Pence (2003), estudia los efectos de las regulaciones estatales de las hipotecas en los límites estatales.

3.6. Métodos de equilibrio general

Cuatro tipos de métodos econométricos de evaluación buscan estimar directamente los efectos del equilibrio general. Primero, como se indicó en las secciones anteriores, los métodos de evaluación estándar de equilibrio parcial pueden tomar los efectos del equilibrio general en los casos en donde la unidad de análisis incorpora estos efectos. Así, por ejemplo, si el tratamiento es asignado aleatoriamente a algunos pueblos y a otros no, y los efectos del equilibrio general ocurren dentro de las poblaciones en lugar de entre ellos, entonces comparar los resultados a nivel de poblaciones, entre poblaciones asignadas aleatoriamente al experimento y poblaciones fuera del experimento, capturarán los efectos del equilibrio general.

El segundo método consiste en los modelos tradicionales (algunos dirían “de estilo antiguo”) de ecuaciones múltiples que relacionan entre sí a varios aspectos del desarrollo económico local. Estos modelos son similares en espíritu a los modelos macroeconómicos de ecuaciones múltiples utilizados por los bancos y otros para generar predicciones económicas de corto plazo, pero con supuestos más específicos de casos de estudio construidos con respecto al contexto particular del desarrollo económico local. Estos modelos han dejado de ser favorecidos en la bibliografía académica porque violan muchas de las reglas de una práctica econométrica, ya que a menudo las ecuaciones consisten en una variable endógena en regresión sobre varias otras, sin instrumentos a la vista. Las bases teóricas precedentes del sistema de ecuaciones múltiples, típicamente carecen por mucho de una estructura formal. La presencia de muchas ecuaciones a menudo dificulta ver de dónde vienen los efectos estimados y, en la práctica, los modelos frecuentemente requieren alguna introducción subjetiva para producir números razonables. En breve, estos modelos son difíciles de ser apreciados, pero la demanda de números en combinación con la falta de alternativas simples los mantiene en juego para las aplicaciones prácticas.

El tercer método es lo que denominé el método del “multiplicador mágico”. Este método desempeña un papel principal en las evaluaciones de los proyectos de tránsito e inversiones de infraestructura deportiva, en donde la firma consultora que lleva a cabo la evaluación, generalmente tiene con anticipación una idea clara del resultado deseado. En este método, se construye alguna medición de los impactos directos de una manera más o menos razonable. Los impactos directos son luego multiplicados por el multiplicador mágico, el cual se supone captura todos los efectos de desplazamiento. Los valores específicos elegidos para el multiplicador mágico, en una aplicación dada, generalmente tienen una escasa justificación formal, lo cual es en sí, uno de sus atractivos para quienes efectúan la evaluación y sus clientes. Ver las investigaciones de Noll y Zimbalist (1997), así como la de Crompton (1995), para más detalles.

El cuarto enfoque, para estimar directamente los efectos del equilibrio general, consiste en especificar un modelo estructural de equilibrio general de la economía relevante que incluye el programa, calibración o estimación de los parámetros del modelo, y luego simular el modelo con y sin el programa.⁶ Los modelos estructurales de equilibrio general hacen suposiciones muy fuertes sobre cómo los diferentes elementos de la economía se afectan unos a otros. Dada la complejidad de estos modelos, el analista no tiene más opción que tratar de manera muy sencilla todo excepto aquellos aspectos de la economía más relevantes para asuntos de interés. Mientras que las fuertes suposiciones y la estructura representan una desventaja en un sentido, también representan las fortalezas de estos modelos. Todos los supuestos son claros y sobre la mesa y definidos, y cualquier efecto que surja del modelo puede ser rastreado hasta aspectos particulares de la economía modelo y por lo tanto regresar a la teoría subyacente. Debido a su complejidad y debido al gasto asociado con estos modelos, emprender el desarrollo de tal modelo tiene sentido sólo para los programas grandes (ya sea en términos de gasto o en términos de la fracción de las unidades relevantes tratadas directamente), donde la respuesta importa mucho y donde esperamos importantes efectos generales de equilibrio.

La bibliografía contiene varias de estas evaluaciones, aquí destaco cuatro ejemplos notables, todos extraídos del campo de las políticas del mercado laboral activo. Davidson y Woodbury (1993) construyen un modelo de búsqueda del equilibrio general que les permite evaluar los efectos del equilibrio general de los programas de bonos de la oficina del Seguro de Desempleo de los Estados Unidos.⁷ Estos programas, cuyos impactos parciales del equilibrio fueron estimados por una serie de experimentos sociales, pagaban a los solicitantes del seguro un bono en efectivo si encontraban trabajo dentro de las primeras 11 semanas del reclamo del beneficio. El modelo de Davidson y Woodbury (1993) indicó que del 30 al 60 por ciento del impacto parcial del equilibrio se cancela en el equilibrio general debido a la sustitución y a los cambios en el esfuerzo de búsqueda óptimo de los

trabajadores desempleados no elegibles para el bono. Lise, Seitz y Smith (2003) modifican el modelo de Davidson y Woodbury (1993) para evaluar el Programa de Autosuficiencia Canadiense, un subsidio al salario para los receptores de asistencia de ingreso de largo plazo que encuentran trabajo de tiempo completo. Ellos encontraron que los efectos de equilibrio generales, incluyendo cambios en el esfuerzo óptimo de búsqueda y en la distribución de los salarios, cambian el desempeño costo-beneficio del programa de positivo a negativo. Blundell, Costa-Dias y Meghir (2003) utilizan los métodos de equilibrio general para examinar un programa de subsidios al salario en el Reino Unido y encuentran que tomando en cuenta los efectos del equilibrio general hay un gran cambio en sus hallazgos. Finalmente, Heckman, Lochner y Taber (1998, 1999) consideran los efectos del equilibrio general en un subsidio a las colegiaturas de 500 dólares anuales para asistir a la universidad en Estados Unidos. Ellos hallaron que tomando en cuenta los cambios en los precios de las herramientas del equilibrio significa que los impactos estimados del equilibrio general son menores que los impactos parciales del equilibrio por un factor de diez. En cada uno de estos casos, tomar en cuenta los efectos del equilibrio general desempeña un papel importante para obtener la respuesta correcta sobre los efectos de un cambio en la política.

3.7. Elección entre métodos de evaluación alternativos no experimentales

Cuando es factible, una evaluación experimental proporcionará la evidencia más precisa sobre la efectividad de los programas locales de desarrollo económico. Cuando un experimento no es factible, el evaluador debe escoger entre los métodos de evaluación alternativos no experimentales resumidos en las Secciones 3.2 a 3.6.

El analista afortunado entra en el proceso al inicio y puede influir el diseño e implementación del programa, así como la recolección de los datos con estimación específica de la estrategia (o estrategias) en mente. En esta feliz situación, el analista puede establecer dos o tres estrategias alternativas de evaluación, proporcionando múltiples líneas de evidencia y permitiendo la (no improbable) posibilidad que una de ellas no se desarrolle en la práctica. El analista desafortunado entra al final del proceso y debe tratar de elegir un método de evaluación que se adecue a los datos e instituciones elegidos por otros.

La bibliografía proporciona bastante dirección para ambos, el analista afortunado y el desafortunado, dirección que frecuentemente es ignorada en la práctica de evaluación. La mayoría de estas direcciones provienen de una creciente lista de artículos que utilizan juegos de datos experimentales para el desempeño *benchmark* (usualmente en términos de prejuicios) de los estimadores de alternativas no experimentales en los diferentes contextos definidos por los

datos disponibles y la organización institucional del programa, donde este último en turno determina la naturaleza del proceso por el cual algunas unidades vienen a recibir tratamiento. Esta bibliografía incluye LaLonde (1986), Fraker y Maynard (1987), Heckman y Hotz (1989), Bell, Orr, Blomquist y Cain (1995), Friedlander y Robins (1995), Heckman, Ichimura y Todd (1997), Heckman, Ichimura, Smith y Todd (1998), Heckman y Smith (1999), Dehejia y Wahba (1999 2002), Agodini y Dynarksi (2001), Glazerman, Levy y Myers (2003), Michalopoulos, Bloom y Hill (2004) y Smith y Todd (2004). La bibliografía también incluye un estudio de simulación, Sección 8.3 de Heckman, LaLonde y Smith (1999) que examina el desempeño de los estimadores alternativos no experimentales, alternativos para varios procesos generadores de datos.

Las primeras partes de esta bibliografía formulan la pregunta de interés en términos de encontrar al único verdadero estimador —la bala mágica que mataría a la bestia de prejuicios de selección en cada contexto. Más recientemente, la bibliografía ha comprendido que ésta no es la pregunta correcta, ya que no existe dicha bala mágica. Como se ha descrito anteriormente, las diferentes estrategias de evaluación no experimentales tienen diferentes suposiciones sobre la naturaleza del proceso de selección y acerca de los datos disponibles. Cuando tales suposiciones se sostienen, un estimador dado proveerá estimaciones consistentes de ciertos parámetros de interés. Y cuando no, no serán consistentes. Así, en lugar de buscar un único estimador que funcione universalmente, la bibliografía busca ahora la planimetría o la relación entre las instituciones y los datos disponibles en un contexto dado (y el parámetro de interés) y la elección de una estrategia de evaluación no experimental. Algunas veces, como en Hui y Smith (2003), los datos disponibles en un contexto dado no apoyan a ningún estimador.

La bibliografía que hace uso de normas de referencia experimentales enseña una cantidad de algunas obvias (aunque no suficientemente obvias, que no han sido ignoradas en muchos artículos publicados e incluso más en los no publicados) pero importantes lecciones. Unos pocos ejemplos sirven para demostrar el valor agregado por esta línea de investigación (Demostración Nacional de Trabajo Apoyado). La evidencia de la serie de artículos que utilizan los datos experimentales de la National Supported Work Demonstration muestra que el conjunto de variables (edad, raza, educación y salario anual demorado) disponible en el Censo de Población Actual de los Estados Unidos no es suficiente para controlar la selección para un programa que sirva a una población con grandes desventajas, incluyendo a ex delincuentes y adictos recuperados. Las estrategias de “selección sobre observables” descrita en la Sección 3.2 requiere datos ricos en las determinantes observables para la participación y los resultados. Heckman, Ichimura, Smith y Todd (1998) demuestran la importancia de tomar miembros de grupos de comparación de los mismos mercados laborales como los participantes cuando se evalúan los programas de mercados laborales activos.

También muestran que utilizar resultados medidos de diferentes maneras para las unidades tratadas y no tratadas (tales como datos administrativos para uno y datos por encuesta para otros) puede llevar a diferencias en los resultados que parezcan un prejuicio de selección pero que en realidad constituyen diferencias sistemáticas de resultados. Heckman y Smith (1999) muestran que cuando los individuos son seleccionados para un tratamiento basado en *shocks* transitorios del mercado laboral, como lo hacen en la mayoría de las políticas activas del mercado laboral, las estrategias de estimadores longitudinales tales como las descritas en la Sección 3.3, las que asumen la selección basada en las diferencias permanentes, se desempeñan de manera bastante pobre. Heckman y Hotz (1989) demuestran el valor de las pruebas de especificación estadística en contextos con datos con la suficiente riqueza para permitir su uso.

Aunque las lecciones específicas de la bibliografía se derivan principalmente de políticas activas del mercado laboral, las lecciones generales se mantienen para la elección de un estimador no experimental para todo tipo de programas de desarrollo económico. Una sana evaluación prestará una detallada atención a la naturaleza de las instituciones que determinan la selección hacia el tratamiento. Estas instituciones determinan la naturaleza de cualquier prejuicio de selección, y por lo tanto la credibilidad de estrategias particulares de evaluación no experimentales. Una sana evaluación prestará detallada atención a la adecuación entre los datos disponibles y el método particular de evaluación utilizado. Los métodos de emparejamiento no tienen sentido sin datos significativos. Los métodos de variables instrumentales no tienen sentido sin un buen instrumento. Los métodos longitudinales no tienen sentido con datos transversales o cuando la selección para el tratamiento depende de características transitorias en lugar de permanentes. En breve, una sana evaluación se desarrolla sobre la teoría económica, la teoría econométrica y la evidencia existente en la elección de una estrategia de evaluación no experimental que empareja los datos y las instituciones presentes en un contexto dado.

4. Alternativas para la evaluación econométrica

4.1. Autoevaluación de los participantes

Los evaluadores podrían ahorrar una gran cantidad de tiempo, dinero y esfuerzo si los participantes del programa pudieran evaluar confiable y directamente al programa. En el caso de un programa que trata a los individuos, esto consiste en pedirles, después de su participación, que determinen si el programa los mejora o no y, si este fuera el caso, ¿qué tanto? Una pregunta sencilla de la encuesta reemplaza a todos los temas econométricos discutidos en la Sección 3. De hecho, muchas evaluaciones de los programas de empleo y

capacitación de los Estados Unidos incluyen dichas preguntas, y los indicadores de desempeño para la Ley de Inversión en la Mano de Obra de los Estados Unidos incluyen medidas de satisfacción del cliente; véase Departamento de Trabajo de los Estados Unidos (2000). Un procedimiento similar podría también aplicarse a las empresas, eligiendo al representante relevante para responder a las preguntas como en Bartik (2004). Los métodos de encuesta incluso podrían ser utilizados para obtener algunos de los efectos del equilibrio general, como se hace en la bibliografía sobre la valuación de los servicios ambientales; ver, Portney (1994). Por ejemplo, se podría preguntar a los residentes locales en cuanto valoran su incubadora para sus nuevos y pequeños negocios, incluso en el caso que ellos no sean usuarios de la misma.

Mientras que la autoevaluación de los participantes suena bien en la teoría, existe sorprendentemente poca evidencia directa en relación a su capacidad para obtener la repuesta correcta. Con el propósito de que las autoevaluaciones de los participantes produzcan estimaciones de impacto válidas, los encuestados tienen que estimar correctamente el contrafactual no observado de lo que les habría sucedido si no hubieran participado, y luego compararlo con la experiencia obtenida como participantes.

Heckman y Smith (1998) presentan algunas evidencias directas sobre la validez de la autoevaluación de los participantes, que reproduzco aquí en la Tabla 12.1. Los miembros del grupo de tratamiento experimental en el Estudio sobre la Ley de Asociación en el Entrenamiento para el Trabajo (JTPA, por sus siglas en inglés) en los Estados Unidos, fueron interrogados en una encuesta de seguimiento, 18 meses después a la asignación aleatoria, si consideraban que el programa los había beneficiado (véanse las notas de la tabla para el parafraseo exacto de las preguntas). La Tabla 12.1 muestra la fracción de cada uno de los cuatro grupos demográficos en el experimento —hombres adultos, mujeres adultas, hombres jóvenes y mujeres jóvenes— quienes respondieron afirmativamente. Los compara con las estimaciones experimentales de impacto sobre las ganancias que cada quien reportó a lo largo del periodo de 18 meses entre la asignación aleatoria y la entrevista de seguimiento. La tabla 12.1 revela poca correlación entre las fracciones que auto reportaron beneficiarse del programa y las estimaciones de impacto experimentales de ingresos, lo que sugiere que los participantes del programa no hacen un buen trabajo en la construcción del contrafactual necesario para determinar si el programa los mejoró o no. Véase Heckman y Smith (1998) o Smith y Whalley (2004) para mayor discusión de la evidencia del experimento sobre la JTPA.

Tabla 12.1 Autoevaluación del impacto de la JTPA: grupo experimental de tratamiento
(Estudio Nacional de la JTPA, muestra de impacto a 18 meses)

	Hombres adultos	Mujeres adultas	Hombres jóvenes	Mujeres jóvenes
	61.63 (0.81)	68.10 (0.68)	62.62 (1.29)	66.29 (1.09)
Porcentajes de las muestras totales				
Porcentajes de quienes auto reportaron su participación				
Porcentaje de participantes auto reportados con autoevaluaciones positivas	62.46 (1.04)	65.21 (0.85)	67.16 (1.59)	71.73 (1.29)
Porcentaje con autoevaluaciones positivas	38.49 (0.81)	44.41 (0.73)	42.06 (1.32)	47.55 (1.16)
<i>Porcentaje de participantes auto reportados con autoevaluaciones positivas por el recibimiento de un tratamiento primario</i>				
Ninguno (abandonos)	48.89 (2.07)	51.44 (1.85)	58.90 (3.33)	61.56 (2.79)
Capacitación en el aula en habilidades ocupacionales	74.10 (2.15)	73.47 (1.36)	72.73 (3.60)	75.28 (2.30)
Capacitación en el trabajo en empresas privadas	75.13 (2.18)	78.90 (2.14)	71.00 (4.56)	75.00 (4.04)
Asistencia en la búsqueda de trabajo	59.57 (2.27)	59.80 (2.18)	68.09 (3.94)	68.94 (4.04)
Educación básica	62.96 (4.67)	56.55 (3.84)	70.97 (4.09)	78.44 (3.19)
Experiencia laboral	66.67 (9.83)	68.75 (5.84)	82.76 (7.14)	73.17 (7.01)
Otros	58.47 (3.65)	66.40 (2.98)	62.50 (4.77)	77.98 (3.99)

Notas:

1. Las proporciones reportadas están basadas en la respuesta a la pregunta “¿Considera que la capacitación u otra asistencia que obtuvo del programa le ayudó a conseguir un trabajo o desempeñarse mejor en el trabajo?”. Esta pregunta fue aplicada solamente a los participantes que se auto reportaron dentro del grupo de tratamiento. La fracción total de autoevaluaciones positivas asume que los no participantes auto reportados habrían provisto autoevaluaciones negativas.
2. El tratamiento primario es en el que el aprendiz haya participado por la mayor cantidad de horas de acuerdo con los registros administrativos de los sitios de JTPA. La mayoría de los aprendices recibieron sólo un servicio; pocos recibieron más de dos. Ver Smith (en impresión) para una exposición detallada. Observe que para algunos participantes que se auto reportaron, los registros administrativos de la JTPA indican que no recibieron ningún servicio.
3. Los errores estándar estimados entre paréntesis.

Fuente: Tabla 8.11 de Heckman y Smith (1998).

La evidencia indirecta también sugiere problemas para los participantes de la autoevaluación. Primero, la discusión en la Sección 3 deja claro que los investigadores tienen gran dificultad en la construcción de estimaciones razonables de los contrafactuales promedio —probablemente una tarea más fácil que estimar el contrafactual de una persona. La bibliografía sobre la teoría de decisiones del comportamiento sugiere que los individuos tienen todo tipo de

problemas cognoscitivos con tareas menos difíciles, como el tomar decisiones consistentes cuando las opciones son presentadas en diferentes formas y que la mayoría de las personas son malos “estadísticos intuitivos”. Los efectos de las encuestas también pueden causar problemas. Los encuestados no querrán correr el riesgo de ofender a un entrevistador al decir que el programa no los ayudó (o no querrán admitirse a sí mismos que gastaron su tiempo y energía en ello). Ver, Bradburn, Sudman y Wansink (2004) en relación con esta clase de entrevistador. Finalmente, como Bartik escribe en su capítulo en este volumen, para los programas que involucran transferencias monetarias (o transferencias valiosas en especie), los entrevistados pueden tener un interés directo en la continuidad del programa y así pueden reportar una respuesta de comportamiento, incluso cuando uno no existe, con el propósito de mantener el flujo de bienes.

Mientras que la limitada evidencia disponible argumenta en contra de confiar en la auto evaluación de los participantes para el análisis de impacto, ésto no impide reunir información útil de los participantes sobre otros aspectos de su participación (o, en verdad, de reunir información útil de los no participantes elegibles acerca del porqué decidieron no tomar parte). Por ejemplo, los participantes probablemente tendrán un buen sentido de la calidad del servicio que recibieron y la cantidad de burocracia involucrada en la participación.

4.2. Estándares de desempeño

Los estándares de desempeño administrativo representan otra alternativa potencial no costosa para el análisis de impacto. Los estándares de desempeño típicamente consisten en indicadores cuantitativos de los rendimientos del programa (número de cheques enviados en tiempo) o resultados (cuantos aprendices encontraron un trabajo dentro del mes posterior a terminar el programa). En los términos de nuestra anotación, generalmente consisten en funciones de Y_1 . Han crecido en popularidad como parte del movimiento de los noventas “reinventando el gobierno” —ver, Osborne y Gaebler (1992)— y ahora invade al gobierno de los Estados Unidos como resultado de la Ley de Desempeño y Resultados Gubernamentales (GPRA, por sus siglas en inglés) de 1993.

Los estándares de desempeño tienen muchos usos y en algunos contextos dicen todo lo que se necesita saber. Por ejemplo, la misión principal de la Administración de la Seguridad Social (SSA, por sus siglas en inglés) de los Estados Unidos consiste en enviar cheques (o efectuar depósitos directos) a las personas indicadas en el tiempo indicado y por el monto correcto. Un indicador del desempeño que muestre la fracción del tiempo en que esto sucede dice mucho (si no es que todo) lo que necesita decirse sobre que tan bien se desempeña la SSA. La tarea de ellos consiste en dar un resultado, en lugar de un impacto, y eso es adecuado para la gerencia que utiliza indicadores de desempeño basados en

resultados. Por supuesto que los cheques que envía la SSA tendrán impactos de comportamiento de interés para los economistas y los diseñadores de políticas; tales impactos no son bien capturados por los indicadores de desempeño basados en niveles de resultados.⁸

En los contextos en que los impactos del programa representan el principal objetivo de preocupación, apoyarse en los estándares de desempeño como aproximación de las estimaciones de impacto requiere de evidencia de una relación sistemática entre ambos. Para concretar el problema, considere un programa de capacitación para el trabajo que utiliza “tasas de ingreso al empleo”, tales como la fracción de personas capacitadas que obtuvieron un empleo a los 90 días de dejar el programa, como su indicador de desempeño. Este indicador es una versión de Y_1 . Con el propósito de que el indicador de desempeño proporcione una aproximación útil del impacto del programa —esto es, de la diferencia que hace en la tasa de empleo con respecto a lo que hubiera ocurrido si los participantes no hubieran participado en el programa—, el indicador de desempeño debe estar correlacionado positivamente con el impacto del programa. En este caso, esto significa que el emplearse después de dejar el programa debe correlacionarse con el cambio en el estatus de empleo inducido por el programa. En un caso extremo, que sería el caso que nadie encontrara trabajo sin el programa, la correlación es igual a uno y el indicador de desempeño es igual al impacto. De manera más general, no hay razón en particular por la cual esta condición deba mantenerse.

Un programa que es más efectivo para aquellos clientes con menos probabilidad de encontrar empleo por su cuenta puede mostrar una relación negativa entre los índices de empleo y los impactos. Para ver esto, considere el caso extremo de clientes fáciles de servir, quienes siempre encontrarían empleo por su cuenta, y clientes difíciles de servir, quienes nunca lo encontrarían pero lo encuentran en la mitad del tiempo cuando participan en el programa. En un mes, el programa sirve a la mitad de cada tipo de clientes. Su tasa de empleo es de 0.75, porque todos los clientes fáciles de servir encuentran empleo y también la mitad de los difíciles de servir, mientras que su impacto es de 0.25, reflejando el hecho que sólo beneficia a los clientes difíciles de servir. En otro mes, sirve sólo a clientes difíciles de servir. En dicho mes, su tasa de empleo es 0.50 pero su impacto también es 0.50, porque ninguno de los clientes difíciles de servir hubiera encontrado empleo por su cuenta. Para este programa, el indicador de desempeño de la tasa de empleo está relacionado negativamente con los impactos del programa, y así proporciona una mala representación para los impactos del programa (y un fuerte incentivo para que los gerentes del programa atiendan sólo a los clientes fáciles de servir).

La evidencia respecto a esta cuestión para los programas de empleo y capacitación de los Estados Unidos, tal como la presentada en Heckman, Heinrich y Smith (2001), y Barnow (1999), y resumida en Barnow y Smith (2003) sugiere

que los indicadores de desempeño comunes utilizados en este contexto, tales como los índices de ingreso al empleo, no se correlacionan muy bien con los impactos del programa.

Situaciones similares aparecen en otros contextos del programa cuando el desempeño es juzgado de acuerdo con los resultados en lugar de los impactos. Por ejemplo, los programas gubernamentales que subsidian la investigación y el desarrollo comercial se juzgan a menudo basados en la fracción de proyectos que reeditúan, lo cual proporciona un incentivo para que los operadores del programa elijan proyectos que habrían obtenido financiamiento de todas formas sin el subsidio, en lugar de financiar proyectos con grandes excedentes (potenciales) pero con pocos beneficios privados. Es este último tipo de proyecto el que proporciona la justificación económica (en oposición a la justificación política) para el subsidio. Ver Wallsten (2000a) para una exposición de divulgación del programa de Investigación de la Innovación de la Pequeña Empresa (SBIR, por sus siglas en inglés) de los Estados Unidos y Wallsten (2000b) para una exposición más académica. La investigación de Wallsten sugiere que cada dólar de los fondos provistos bajo el programa SBIR se excluye un dólar de recursos privados de investigación. Esto ocurre como resultado de un enfoque para elegir proyectos con probabilidades de tener éxito en el mercado, ya que ésta es la métrica utilizada para juzgar el éxito del programa. Sus hallazgos indican que el programa no tiene impacto sobre el monto total de la investigación emprendida. En conjunto, la bibliografía sugiere que los estándares de desempeño no representan un sustituto general para la evaluación econométrica de impacto.

5. La práctica

Esta sección considera brevemente algunos temas prácticos importantes asociados con la evaluación de las políticas locales para el desarrollo económico. La primera subsección se enfoca en cuándo no realizar una evaluación. La segunda subsección se enfoca en cómo seleccionar un evaluador y luego cómo evaluar una evaluación una vez terminada, y el tercero resalta algunos temas importantes en el análisis costo-beneficio. Esta sección se desarrolla en lo expuesto en Smith y Sweetman (2001).

5.1. Cuando no hacer una evaluación

Las evaluaciones consumen tiempo y recursos. Por lo tanto, las evaluaciones así como los programas de evaluación, deben emprenderse sólo cuando sus beneficios tienen la probabilidad de exceder a sus costos. Esta subsección esboza un número de situaciones en las cuales la evaluación no tiene probabilidad de pasar una prueba estándar de costo-beneficio y en las cuales, como resultado, el dinero que sería gastado en la evaluación sería mejor destinarlo a otras cosas.

La primera situación en la cual una evaluación es una mala idea es cuando el dinero escasea y las otras funciones administrativas básicas no están en orden. Antes de realizar una evaluación, los operadores del programa deberían tener una idea clara de cuáles unidades participan en su programa, determinar si las unidades participantes son elegibles o no para el programa y, en un programa voluntario, cómo se comparan las unidades participantes con la población de todas las unidades elegibles. También deberían saber cuánto dinero está gastando el programa, en qué se está gastando y en cuáles de las unidades tratadas está siendo gastado el dinero. Recolectar y examinar toda esta información representa una obligación fiduciaria básica por parte de los operadores del programa, actuando como agentes de los siempre sufridos contribuyentes de impuestos (o para aquellos quienes hacen donaciones a la organización sin fines de lucro que patrocina los proyectos de desarrollo económico). Estas labores fiduciarias deberían darse antes de intentar una evaluación; después de todo, un programa que no está bajo control en términos de elegibilidad y costos parece poco probable que produzca mucho a manera de impacto.

La segunda situación donde se puede evitar la evaluación es cuando el impacto se conoce con anticipación. Esto puede suceder por dos razones. La primera, puede ser que ya exista bibliografía de evaluación sustancial y de alta calidad para un tipo de programa particular. Por ejemplo, gracias a una larga serie de evaluaciones experimentales a nivel estatal, la bibliografía tiene una idea bastante clara sobre los efectos de los programas de asistencia en la búsqueda de trabajo, tanto voluntarios como obligatorios, para las madres solteras sobre el bienestar en los Estados Unidos. Véase Gueron y Pauly (1991), y Bloom y Michalopoulos (2001). Las evaluaciones adicionales, a menos que cubran una modalidad de tratamiento no trivial o población diferente, probablemente no justifican su costo. La segunda forma para que la respuesta se conozca por anticipado es cuando los programas realmente existen sólo para transferir dinero a algunos individuos, empresas u otros intereses favorecidos, con la justificación del desarrollo económico, sirviendo como una distracción útil para el público apático y los medios ineptos. Los subsidios para empresas particularmente grandes en busca de la ubicación de una nueva planta representan un ejemplo importante de tales programas. Desde una perspectiva nacional, tales guerras de ofertas entre los estados y las localidades no pueden hacer más que tener un impacto cero, y tendrán uno negativo en la medida que sean la causa de asignaciones geográficas inadecuadas de la producción. Así, la Unión Europea prohíbe las competencias de este tipo entre sus naciones miembro.⁹ Estos programas también socavan los preceptos de la ley, una determinante importante del crecimiento a largo plazo a nivel internacional, al tratar diferente a algunas empresas tan sólo por su tamaño, movilidad o conexiones políticas. Un poco de cinismo, combinado con una sencilla teoría económica y una atención cuidadosa hacia donde se dirige el dinero, generalmente es suficiente para identificar dichas

situaciones. Tristemente, exponerlas a la luz del día no siempre (o incluso a menudo) causa su desaparición.

La tercera situación en la cual una evaluación puede no representar una buena inversión surge cuando las muestras disponibles para la evaluación carecen del tamaño requerido para la deducción estadística. Por ejemplo, un programa que subsidia únicamente a cinco empresas no proporcionará una visión estadística clara de los impactos del programa utilizando cualquiera de los métodos perfilados en las Secciones 3.1 a 3.6. Los métodos econométricos esbozados aquí requieren un número sustancial de unidades tratadas y no tratadas con el propósito de identificar los impactos estadísticamente distinguibles de cero. Los tamaños limitados de las muestras también pueden surgir debido a las limitaciones presupuestales en situaciones con un gran número de unidades potencialmente disponibles para una evaluación, pero con costos sustanciales (ver los costos de encuestas) por agregar cada unidad a los datos. Los estadísticos han desarrollado métodos formales para determinar el número de observaciones requeridas para detectar un impacto de un tamaño dado con un cierto grado de confianza bajo varios supuestos sobre la variante de los indicadores de resultados y otras características del contexto de la evaluación. La aplicación de estos métodos constituye un “análisis del poder”, dicho análisis debería preceder la decisión de determinar si se lleva a cabo una evaluación, excepto en los casos de tamaños de muestra muy pequeños o muy grandes. Véase, Maxwell (2000) para una exposición del análisis de poder y citas adicionales.

La cuarta situación que requiere no realizar una evaluación surge cuando no existen datos para apoyar la evaluación o cuando no se pueden obtener datos de alta calidad a un costo dentro del presupuesto disponible. Por ejemplo, muchas evaluaciones relativamente caras de los principales programas gubernamentales a menudo se apoyan en datos de encuesta con índices de respuesta asombrosamente (o incluso espantosamente) bajos. En tales situaciones, la evaluación debe dedicar atención adicional a temas de respuesta no selectiva, que pueden arrojar dudas sobre la consistencia y generalidad de los hallazgos. En tales situaciones, podría ser mejor el invertir suficiente dinero para obtener un índice de respuesta razonable (digamos un 80 por ciento) o utilizar fuentes alternativas de datos, como los datos administrativos. Los recientes por supuesto, no son una panacea. La calidad de los datos tiene una baja prioridad en muchas agencias, lo cual significa que una evaluación debe dedicar recursos sustanciales a la limpieza después de contabilizar los datos, quizás descartando por completo algunos campos. Véase Hotz y Scholz (2002) para una mayor exposición sobre los datos administrativos. En general, existe un nivel de reserva de la calidad de los datos por debajo de la cual una evaluación pierde todo valor.

La falta de experiencia para realizar la evaluación constituye la situación final cuando ninguna evaluación puede dominar cualquier otra evaluación. Si la

organización que lleva a cabo potencialmente una evaluación carece del acceso a los fondos o del personal necesario para desarrollarla, o de los medios para evaluarla una vez terminada, entonces por lo general no se debería intentar realizar la evaluación, particularmente si existen evaluaciones relativamente fuertes en programas similares operados en contextos similares. Las evaluaciones débiles no justifican el dinero que cuestan, lo cual conduce directamente al tema de la siguiente subsección, el cual se ocupa de cómo seleccionar un evaluador y cómo evaluar la evaluación del evaluador.

5.2. Elección y evaluación de un evaluador

Cualquiera puede llamarse a sí mismo un evaluador y buscar contratos para desempeñar evaluaciones. En la práctica, muchos individuos (y conjuntos de individuos en empresas o coaliciones para proyectos específicos) llevan a cabo evaluaciones, incluyendo economistas, estadísticos, psicólogos y sociólogos. Algunos realizan evaluaciones lateralmente, además de enseñar o hacer investigaciones académicas; otros hacen evaluaciones de tiempo completo. Algunos conocen lo último en métodos econométricos mientras otros sólo pueden hacer corridas de regresiones. El contexto particular de evaluación y el presupuesto cobran mucha importancia aquí, por lo que sólo ofrezco unas cuantas observaciones generales.

Primero, los experimentos son más difíciles de lo que piensan; si desean realizar uno y nunca antes lo han hecho, contraten a una firma que sepa como hacerlo. Dichas empresas incluyen a MDRC, Matemática y Abt Associates, por nombrar a algunas. Segundo, diferentes tipos de evaluadores tienen diferentes características, las cuales deben corresponder con las necesidades actuales. Las empresas de evaluación profesional son las más caras, pero tienen mucha experiencia y entregan un producto bien pulido, a tiempo y, generalmente, dentro del presupuesto. Por otro lado, los académicos tienden a costar un poco menos y algunas veces saben más de econometría, pero tienen una menor probabilidad de concluir a tiempo y en forma menos pulida.

Tercero, algunos evaluadores tomarán el dinero y devolverán un revoltijo embarazoso. Por ejemplo, el artículo de Gregory (2000), publicado (por una razón desconocida) en el diario *Evaluation*, sugiere una evaluación centrada en una variante de la “matriz de sitios de opresión”. En este enfoque de evaluación, las partes interesadas clave se sientan a reflexionar en todas las maneras en las cuales la vida los ha tratado mal (o al menos aquellas maneras, de alguna forma, relacionadas con el programa), y el evaluador escribe sobre la matriz a llenar. Es bastante obvio, que dichos ejercicios representan una entretenida diversión para las partes interesadas y dinero fácil para el evaluador, pero no rinde discernimiento alguno sobre el impacto del programa. Cuarto, algunas veces se puede conseguir

que la parte econométrica de una evaluación se elabore a bajo costo, si a un investigador académico se le proporcionan datos interesantes que pueda utilizar para escribir artículos para publicaciones en diarios académicos. Verdaderamente muchas, sino la mayoría, de las evaluaciones de programas sociales publicadas no fueron pagadas por las agencias que operan el programa que evalúan.

Las evaluaciones, como los programas, requieren, a su vez, de evaluación. Algunas evaluaciones son muy buenas mientras que otras son muy débiles. No todas las agencias que comisionan evaluaciones tienen la experiencia con el personal interno para emprender evaluaciones. Aun cuando la tienen, las revisiones externas de calidad pueden agregar valor sustancial a las evaluaciones y también incrementar su credibilidad. Existen una variedad de métodos para incorporar retroalimentación externa y revisar el proceso de evaluación. Las evaluaciones de gran escala a menudo incluyen un panel de revisión técnico compuesto por expertos quienes proporcionan retroalimentación en etapas críticas, tales como el reporte de diseño y el borrador del análisis de impacto. En evaluaciones menores, un solo experto externo puede realizar este papel proporcionando comentarios sobre los borradores de varios reportes. Una vez que la evaluación está completa, la retroalimentación es todavía útil para guiar a los lectores en la determinación de cuánto peso asignar a los resultados obtenidos. Alentar la publicación en diarios determinados es una forma de lograr esto; y el saber que el producto final será enviado para arbitraje, proporciona un incentivo para la calidad a través del proceso de evaluación. La inclusión de comentarios por un revisor, como apéndice al reporte final publicado, como en Jacobson y Petta (2000), desempeña un papel similar.

5.3. Análisis costo-beneficio

La política basada en evidencia se construye sobre el cimiento de un análisis costo-beneficio serio y profundo de varias alternativas de la política. Los análisis costo-beneficio descansan sobre una base de evaluaciones econométricas de alta calidad del programa. Muchos árboles han prestado sus vidas para libros sobre la evaluación costo-beneficio. Esta subsección no intenta un tratamiento general, sino que destaca algunos aspectos clave que han recibido muy poca atención en la bibliografía. La exposición aquí retoma parte de la discusión en la Sección 10.2 de Heckman, LaLonde y Smith (1999).

Las evaluaciones de impacto comúnmente generan estimaciones del impacto neto por un tiempo corto, usualmente unos cuantos meses o años. Todavía, al menos en ciertos contextos, tales como los programas de desarrollo del capital humano o las inversiones en infraestructura, esperamos que los impactos persistan por algún tiempo. En tales contextos, surgen dos temas relacionados. El primero es cómo proyectar los beneficios estimados fuera del periodo de los

datos. Mientras que la teoría o la evidencia de otras evaluaciones de programas similares con plazos de seguimiento más largos, pueda desempeñar un papel para guiar esta decisión, el mejor curso de acción probablemente consistirá al final en construir estimaciones del desempeño del costo-beneficio del programa, asumiendo múltiples duraciones posibles para los beneficios del mismo. La reciente evaluación experimental de los Cuerpos de Trabajo no hace esto y como resultado, particularmente en su resumen ejecutivo, proporciona una guía algo engañosa para la política. Dicha evaluación presume como punto de partida que los impactos del programa durarán (esencialmente) para siempre y concluye con base en eso que el programa supera fácilmente una prueba costo-beneficio [véase la Tabla 3 en Burghardt et al. (2001) y la discusión respectiva]. Sin embargo, debido al alto costo de este programa y el periodo relativamente corto para la recolección de datos después del programa, sin la suposición de los beneficios futuros, el programa hubiera tenido un impacto bruto positivo (lo cual representa uno de los principales logros en relación con la mayoría de los programas gubernamentales de trabajo y entrenamiento para la juventud) pero hubiera fallado miserablemente la prueba costo-beneficio.

El segundo tema es qué tasa de descuento se debe utilizar para los beneficios futuros. Existe poca bibliografía que intenta estimar las tasas de descuento sociales, óptimas bajo varios supuestos [véase, la exposición y referencias en Liu (2003)]. Una vez más, el reportar estimaciones de costo-beneficio para varias tasas posibles parece ser mejor.

Otro tema en el análisis costo-beneficio se interesa por el costo de peso muerto de la tributación, llamado “obligación excesiva” en la bibliografía de finanzas públicas. Esta cantidad mide el costo en la economía de un dólar de impuesto marginal, incluyendo (idealmente) tanto los costos directos de operar el sistema fiscal como los costos indirectos de las distorsiones inducidas por el sistema fiscal. La bibliografía ofrece una amplia variedad de estimaciones de estos costos, que van desde unos pocos centavos hasta más de un dólar por cada dólar de ingresos fiscales. Véase Browning (1987) y Snow y Warren (1996) para una mayor exposición y evidencia. Una vez más, dada la incertidumbre en la bibliografía, presentar múltiples estimaciones basadas en valores diferentes parece ser el mejor rumbo (e ignorar los costos de peso muerto en su conjunto, como lo hacen muchas evaluaciones, parece ser el peor camino).

Para el caso de cada tema de costo-beneficio considerado aquí, el presentar estimaciones múltiples que se apoyen en diferentes suposiciones, hace dos cosas importantes. Primero, permite a los lectores con diferentes opiniones, a priori a las del evaluador sobre estos temas, ver las estimaciones de costo-beneficio bajo sus suposiciones preferidas. Segundo, subraya el rango de estimaciones de costo-beneficio consistentes con los datos para los diseñadores de políticas. Esto los obliga a enfrentar la incertidumbre acerca del desempeño del programa y,

al mismo tiempo, proporciona un sentido robusto de cualquier recomendación referente a la política.

6. Conclusión

Los recursos comprometidos con los programas de desarrollo económico local tienen usos alternativos valiosos. Las evaluaciones econométricas de los programas juegan un papel clave para determinar cuando continuar con los programas de desarrollo económico y cuando terminarlos. En este capítulo he enfatizado cinco temas clave para la evaluación de tales programas.

Primero, y quizás antes que nada, he enfatizado la importancia de leer la bibliografía. Hemos aprendido mucho acerca de cómo hacer las evaluaciones de políticas econométricas en las dos últimas décadas, sin embargo este conocimiento aún no ha afectado la práctica de la evaluación en la medida que debería hacerlo. El conocimiento que hemos obtenido incluye tanto avances en los métodos como avances en la práctica, incluyendo el uso de datos administrativos y estrategias inteligentes de identificación.

Segundo, no existe una bala mágica. Ningún estimador para la evaluación econométrica proporciona estimaciones consistentes para todas (o incluso la mayoría) las posibles combinaciones de datos, instituciones y parámetros de interés. La regresión no hace esto, la diferencia en las diferencias no lo hace y el VI no lo hace. La búsqueda de dicho estimador, que impulsó la bibliografía por muchos años, ha llegado a su fin, reemplazada por un programa de investigación más sensible diseñado para identificar la planimetría entre las características de los datos y las instituciones, y los parámetros de interés con los estimadores que probablemente produzcan respuestas consistentes.

Tercero, los efectos heterogéneos de tratamiento importan. Afectan la elección e interpretación de los estimadores econométricos para la evaluación. Implican un pensamiento cuidadoso sobre el parámetro de interés exacto que se requiere para responder una pregunta particular de la política. La bibliografía conceptual está avanzando rápidamente en esta área y ya ha revolucionado el pensamiento de los evaluadores sobre lo que hacen.

Cuarto, los efectos del equilibrio general son importantes en muchos contextos de evaluación, particularmente al considerar los programas para el desarrollo económico local, los cuales generalmente están dirigidos a crear dichos efectos. La presencia de la potencia de los efectos del equilibrio general tienen importantes implicaciones para la decisión conjunta, en relación con cual estimador de evaluación econométrica emplear y la unidad del análisis para la evaluación. Dependiendo de la unidad del análisis, algunos estimadores fallarán o, aún peor, serán sesgados por los efectos del equilibrio general. Elija la unidad

de análisis muy alta y los efectos del programa se perderán entre las capas; elija la unidad de análisis muy baja y los efectos del equilibrio general se perderán. Así mismo, parece ser que los efectos de equilibrio general atraen malas prácticas de evaluación, en particular el uso de multiplicadores mágicos en la evaluación de la inversión pública para la infraestructura.

Finalmente, no todo programa se beneficiará de una evaluación. Antes de proceder con una reflexión, realice algunos cálculos sobre el poder explicativo y de representatividad, y haga un análisis informal del costo-beneficio de la evaluación, esto le ayudará a separar las situaciones en las cuales una evaluación representa una buena inversión de las situaciones en las que representa una pérdida de tiempo y dinero.

Notas

1. Por supuesto, la suma de algunas habilidades técnicas no sólo representará un desarrollo personal en la parte de desarrolladores económicos de filiación contraria, sino que también puede mejorar su habilidad de promover el desarrollo económico local.
2. Por ejemplo, los reportes de efectos marginales de Stata son equivalentes a las derivadas evaluadas en la media de covariaciones, en oposición a las ligeramente más difíciles, pero técnicamente preferidas para calcular la derivada media al tomar el promedio de derivadas evaluadas en los valores de covariaciones para cada observación.
3. En la bibliografía técnica, esto es llamado “el curso de dimensionalidad”. Ver Smith y Todd (2004) para una discusión más amplia.
4. La bibliografía que utiliza los modelos de panel para evaluar el impacto del cambio de Ayuda a las Familias con Hijos Dependientes a Ayuda Temporal para Familias Necesitadas en los Estados Unidos proporciona un buen ejemplo de los peligros de utilizar estos modelos con cada pequeña variación al momento del tratamiento. Esta bibliografía se apoya en una variación mensual limitada en la implementación del TANF (por sus siglas en inglés) a través de los estados – variación que parece estar relacionada con los resultados bajo estudio y por lo tanto viola las suposiciones que justifican los modelos de panel.
5. El modelo también puede ser estimado en una etapa utilizando toda la información de los métodos de probabilidad máxima. Mientras exista una ganancia de eficiencia al hacerlo, la versión de dos etapas puede ser más robusta o con falta de especificaciones, ya que recae con menos fortaleza que la suposición de normalidad.
6. El debate relacionado con los méritos relativos de calibración y estimación de los modelos de equilibrio estructural yace más allá del alcance de este documento. Ver Hansen y Heckman (1996), Kyland y Prescott (1996), y Sims (1996) para tres presentaciones relativas no técnicas de diferentes puntos de vista del debate.
7. Ver Meyer (1995) para un repaso de los experimentos de política relacionados con el Seguro de Desempleo.
8. He tomado este ejemplo de Wilson (2000), un libro valioso de leer para aquellos interesados en por qué la administración y evaluación de los programas gubernamentales es difícil de comprobar en la práctica.
9. Ver la discusión y las referencias en Bartik (2003b).

Referencias bibliográficas

- AGODINI, Roberto y Mark DYNARSKI (2001), "Are Experiments the Only Option? A Look at Dropout Prevention Programs". *Mathematica Policy Research Working Paper* núm. 8723-300.
- ANGRIST, Joshua y Guido IMBENS (1995), "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity". *Journal of the American Statistical Association*. 90(430): 431-432.
- ANGRIST, Joshua y Alan KRUEGER (1999), "Empirical Strategies in Labor Economics". En Orley Ashenfelter y David Card, eds., *Handbook of Labor Economics*, volumen 3A. Amsterdam: Holanda del Norte. 1277-1366.
- ANGRIST, Joshua y Alan KRUEGER (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments". *Journal of Economic Perspectives*. 15(4): 69-85.
- AYRES, Ian y John DONOHUE (2002), "Shooting Down the More Guns, Less Crime Hypothesis". NBER Working Paper núm. 9336.
- BARNOW, Burt (1999), "Exploring the Relationship between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act". *Journal of Policy Analysis and Management*. 19(1): 118-141.
- BARNOW, Burt y Jeffrey SMITH (2003), "Performance Management of US Job Training Programs". Manuscrito inédito, University of Maryland.
- BARTIK, Timothy (2003a), "Local Economic Development Policies". W.E. Upjohn Institute Staff Working Paper núm. 03-91.
- BARTIK, Timothy (2003b), "Thoughts on American Manufacturing Decline and Revitalization". *Employment Research*. 10(4): 1-4.
- BARTIK, Timothy (2004), "Evaluating the Impacts of Local Economic Development Policies On Local Economic Outcomes: What Has Been Done and What is Doable?" En este libro.
- BELL, Stephen, Larry ORR, John BLOMQUIST y Glen CAIN (1995), *Program Applicants as a Comparison Group in Evaluating Training Programs*. Kalamazoo: W.E. Upjohn Institute for Employment Research.
- BLACK, Dan y Daniel NAGIN (1998), "Do Right-to-Carry Laws Deter Violent Crime?" *Journal of Legal Studies*. 27(1): 209-219.
- BLACK, Dan y Jeffrey SMITH (2004), "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching". *Journal of Econometrics*, por publicarse.
- BLACK, Dan, Jeffrey SMITH, Mark BERGER y Brett NOEL (2003), "Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System". *American Economic Review*. 93(4): 1313-1327.
- BLACK, Sandra (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education". *Quarterly Journal of Economics*. 114(2): 577-599.
- BLOOM, Dan y Charles MICHALOPOULOS (2001), *How Welfare and Work Policies Affect Employment and Income: A Synthesis of Research*. New York: Manpower Demonstration Research Corporation.

- BLOOM, Howard (1984), "Accounting for No-Shows in Experimental Evaluation Designs". *Evaluation Review*. 82(2): 225-246.
- BLUNDELL, Richard y Monica COSTA DIAS (2000), "Evaluation Methods for Non-Experimental Data". *Fiscal Studies*. 21(4): 427-468.
- BLUNDELL, Richard y Monica COSTA DIAS (2002), "Alternative Approaches to Evaluation in Empirical Microeconomics". *Portuguese Economic Journal*. 1(1): 91-115.
- BLUNDELL, Richard, Monica COSTA DIAS y Costas MEGHIR (2003), "The Impact of Wage Subsidies: A General Equilibrium Approach". Texto inédito, University College, Londres.
- BLUNDELL, Richard, Monica COSTA DIAS, Costas MEGHIR y John VAN REENAN (2001), "Evaluating the Impact of a Mandatory Job Search Assistance Program". IFS Working Paper núm. WP01/20.
- BLUNDELL, Richard y James POWEL (2001), "Endogeneity in Semiparametric Binary Response Models". *CEMMAP Working Paper*, núm. CWPO5/01.
- BRADBURN, Norman, Seymour SUDMAN and Brian WANSINK (2004), *Asking Questions: The Definitive Guide to Questionnaire Design*. Jossey-Bass.
- BROWNING, Edgar (1987), "On the Marginal Welfare Cost of Taxation". *American Economic Review*. 77(1): 11-23.
- BURGHARDT, John, Peter SCHOCHET, Sheena MCCONNELL, Terry JOHNSON, Mark GRITZ, Steven GLAZERMAN, John HOMRIGHAUSEN y Russell JACKSON (2001), *Does the Job Corps Work? Summary of the National Job Corps Study*. Princeton, NJ: Mathematica Policy Research.
- CALMFORS, Lars (1994), "Active Labor Market Policy and Unemployment – A Framework for the Analysis of Crucial Design Features". *OCDE Economic Studies*. 22(1): 7-47.
- CARD, David (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". En Louis Christofides, Kenneth Grant y Robert Swidinsky, eds., *Aspects of Labour Market Behavior: Essays in Honor of John Vanderkamp*. Toronto: University of Toronto Press. 201-222.
- COATES, Dennis y Brad HUMPHREYS (1999), "The Growth Effects of Sport Franchises, Stadia and Arenas". *Journal of Policy Analysis and Management*. 18(4): 601-624.
- CROMPTON, John (1995), "Analysis of Sports Facilities and Events: Eleven Sources of Misapplication". *Journal of Sports Management*. 9(1): 14-35.
- DAVIDSON, Carl y Stephen WOODBURY (1993), "The Displacement Effects of Reemployment Bonus Programs". *Journal of Labor Economics*. 11(4): 575-605.
- DEHEJIA, Rajeev y Sadek WAHBA (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs". *Journal of the American Statistical Association*. 94(448): 1053-1062.
- DEHEJIA, Rajeev y Sadek WAHBA (2002), "Propensity Score Matching Methods for Non-Experimental Causal Studies". *Review of Economics and Statistics*. 84(1): 151-161.
- DOOLITTLE, Frederick y Linda TRAEGER (1990), *Implementing the National JTPA Study*. Nueva York: Manpower Demonstration Research Corporation.

- EICHLER, Martin y Michael LECHNER (2002), "An Evaluation of Public Employment Programmes in the East German State of Sachsen-Anhalt". *Labour Economics*. 9(2): 143-186.
- EVANS, William y Diana LIEN (2002), "The Benefits of Prenatal Care: Evidence from the PAT Bus Strike". Texto inédito, University of Maryland.
- EVANS, William y Julie TOPOLESKI (2002), "The Social and Economic Impact of Native American Casinos". NBER Working Paper núm. 9198.
- EBERTS, Randall y Christopher O'LEARY (2004), "Evaluating Training Programs: Impacts at the Local Level". En este libro.
- FRAKER, Thomas y Rebecca MAYNARD (1987), "The Adequacy of Comparison Group Designs for Evaluation of Employment-Related Programs". *Journal of Human Resources*. 22(2): 194-227.
- FRIEDLANDER, Daniel y Philip ROBINS (1995), "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods". *American Economic Review*. 85(4): 923-937.
- FRÖLICH, Markus (2004), "Finite Sample Properties of Propensity Score Matching and Weighting Estimators". *Review of Economics and Statistics*. Por publicarse.
- GLAZERMAN, Steven, Dan LEVY y David MYERS (2003), "Nonexperimental Versus Experimental Estimates of Earnings Impacts". *Annals of the American Academy of Political and Social Science*. 589: 63-93.
- GREENBERG, David y Mark SHRODER (1997), *Digest of Social Experiments*, 2ª edición. Lanham, Maryland: Rowman and Littlefield.
- GREENE, William (2002), *Econometric Analysis*, 5a edición. Upper Saddle River, NJ: Prentice Hall.
- GREGORY, Amanda (2000), "Problematizing Participation: A Critical Review of Approaches to Participation in Evaluation Theory". *Evaluation*. 6(2): 179-199.
- GUERON, Judith y Edward PAULY (1991), *From Welfare to Work*. Nueva York: Russell Sage.
- HAHN, Jinyong (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects". *Econometrica*. 66(2): 315-331.
- HAHN, Jinyong, Petra TODD y Wilbert VAN DER KLAUW (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design". *Econometrica*. 69(1): 201-209.
- HANSEN, Lars y James HECKMAN (1996), "The Empirical Foundations of Calibration". *Journal of Economic Perspectives*. 10(1): 87-104.
- HECKMAN, James (1979), "Sample Selection Bias as a Specification Error". *Econometrica*. 47(1): 153-161.
- HECKMAN, James (1996), "Randomization as an Instrumental Variable". *Review of Economics and Statistics*. 78(2): 336-341.
- HECKMAN, James, Carolyn HEINRICH y Jeffrey SMITH (2001), "The Performance of Performance Standards". *Journal of Human Resources*. 37(4): 778-811.

- HECKMAN, James, Neil HOHMANN, Jeffrey SMITH y Michael KHOO (2000), "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment". *Quarterly Journal of Economics*. 115(2): 651-694.
- HECKMAN, James y V. Joseph HOTZ (1989), "Choosing Among Alternative Methods of Estimating the Impact of Social Programs: The Case of Manpower Training". *Journal of the American Statistical Association*. 84(408): 862-874.
- HECKMAN, James, Hidehiko ICHIMURA, Jeffrey SMITH y Petra TODD (1998), "Characterizing Selection Bias Using Experimental Data". *Econometrica*. 66(5): 1017-1098.
- HECKMAN, James, Hidehiko ICHIMURA y Petra TODD (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme". *Review of Economic Studies*. 64(4): 605-654.
- HECKMAN, James, Hidehiko ICHIMURA y Petra TODD (1998), "Matching as an Econometric Evaluation Estimator". *Review of Economic Studies*. 65(2): 261-294.
- HECKMAN, James, Robert LALONDE y Jeffrey SMITH (1999), "The Economics and Econometrics of Active Labor Market Programs". En Orley Ashenfelter y David Card, eds., *Handbook of Labor Economics*, volumen 3A. Amsterdam: North Holland. 1865-2097.
- HECKMAN, James, Lance LOCHNER y Christopher TABER (1998), "General Equilibrium Treatment Effects: A Study of Tuition Policy". *American Economic Review*. 88(2): 281-386.
- HECKMAN, James, Lance LOCHNER y Christopher TABER (1999), "General Equilibrium Cost-Benefit Analysis of Education and Tax Policies". En Gustav Ranis y Kakshmi Raut, eds., *Taxes, Growth and Development: Essays in Honor of Professor T.N. Srinivasan*. New York: Elsevier. 291-349.
- HECKMAN, James y Salvador NAVARRO-LOZANO (2004), "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models". *Review of Economics and Statistics*. Por publicarse.
- HECKMAN, James y Richard ROBB (1985), "Alternative Methods for Evaluating the Impact of Interventions". En James Heckman y Burton Singer, eds., *Longitudinal Analysis of Labor Market Data*. Nueva York: Cambridge University Press para la Serie de Monografías de la Econometric Society. 156-246.
- HECKMAN, James y Jeffrey SMITH (1995), "Assessing the Case for Social Experiments". *Journal of Economic Perspectives*. 9(2): 85-110.
- HECKMAN, James y Jeffrey SMITH (1998), "Evaluating the Welfare State". En Steiner Strom, ed., *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*. Cambridge University Press para la Serie de Monografías de la Econometric Society. 241-318.
- HECKMAN, James y Jeffrey SMITH (1999), "The Pre-Programme Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies". *Economic Journal*. 109(457): 313-348.
- HECKMAN, James, Jeffrey SMITH y Nancy CLEMENTS (1997), "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Program Impacts". *Review of Economic Studies*. 64(4): 487-537.

- HECKMAN, James, Jeffrey SMITH y Christopher TABER (1998), "Accounting for Dropouts in Evaluations of Social Programs". *Review of Economics and Statistics*. 80(1): 1-14.
- HECKMAN, James, Justin TOBIAS y Edward VYTLACIL (2003), "Simple Estimators for Treatment Parameters in a Latent Variable Framework". *Review of Economics and Statistics*. 85(3): 748-754.
- HECKMAN, James y Edward VYTLACIL (1998), "Instrumental Variable Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling when the Return Is Correlated with Schooling". *Journal of Human Resources*. 33(4): 974-987.
- HECKMAN, James y Edward VYTLACIL (2001a), "Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return to Schooling". *Review of Economics and Statistics*. 83(1): 1-12.
- HECKMAN, James y Edward VYTLACIL (2001b), "Local Instrumental Variables". En Cheng Hsiao, Kimio Morimune y James Powell, eds., *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*. Cambridge: Cambridge University Press.
- HECKMAN, James y Edward VYTLACIL (2004), "The Econometric Evaluation of Social Programs". En James Heckman y Edward Leamer, eds. *Handbook of Econometrics*, Volume 6. Amsterdam: North Holland. Por publicarse.
- HOLLISTER, Robinson, Peter KEMPER y Rebecca MAYNARD (1984), *The National Supported Work Demonstration Project*. Madison: University of Wisconsin Press.
- HOTZ, V. Joseph y Karl SCHOLZ (2002), "Measuring Employment and Income Outcomes for Low-Income Populations with Administrative and Survey Data". En *Studies of Welfare Populations: Data Collection and Research Issues*. National Research Council: National Academy Press. 275-315.
- HUI, Shek-Wai y J. SMITH (2002), "The Labor Market Impacts of Adult Education and Training in Canada". Reporte preparado para Human Resources Development Canadá.
- IMBENS, Guido (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions". *Biometrika*. 87(3): 706-710.
- IMBENS, Guido (2004), "Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review". *Review of Economics and Statistics*, por publicarse.
- IMBENS, Guido y Joshua ANGRIST (1994), "Identification and Estimation of Local Average Treatment Effects". *Econometrica*. 62(4): 467-476.
- JACOBSON, Lou e Ian PETTA (2000), "Measuring the Effects of Public Labor Exchange (PLX) Referrals and Placements in Washington and Oregon". *Workforce Security Occasional Paper* núm. 2000-06, Employment and Training Administration, Departamento del Trabajo de los EU.
- KYDLAND, Finn y Edward PRESCOTT (1996), "The Computational Experiment: An Econometric Tool". *Journal of Economic Perspectives*. 10(1): 69-85.
- LALONDE, Robert (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data". *American Economic Review*. 76(4): 604-620.

- LECHNER, Michael (1999), "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany After Unification". *Journal of Business and Economic Statistics*. 17(1): 74-90.
- LECHNER, Michael (2000), "An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany". *Journal of Human Resources*. 35(2): 347-375.
- LECHNER, Michael (2001), "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption". En Michael Lechner y Friedhelm Pfeiffer, eds., *Econometric Evaluation of Labor Market Policies*. Heidelberg: Physica. 43-58.
- LIU, Liqun (2003), "A Marginal Cost of Funds Approach to Multi-Period Public Project Evaluation: Implications for the Social Discount Rate". *Journal of Public Economics*. 87(7-8): 1707-1718.
- LISE, Jeremy, Shannon SEITZ y Jeffrey SMITH (2003), "Equilibrium Policy Experiments and the Evaluation of Social Programs". Texto inédito, University of Maryland.
- LONG, Sharon y Douglas WISSOKER (1995), "Welfare Reform at Three Years: The Case of Washington's Family Independence Program". *Journal of Human Resources*. 30(4): 766-790.
- LOTT, John y David MUSTARD (1997), "Crime, Deterrence and Right-to-Carry Concealed Handguns". *Journal of Legal Studies*. 26(1): 1-68.
- MAXWELL, Scott (2000), "Sample Size and Multiple Regression Analysis". *Psychological Methods*. 5(4): 434-458.
- MEYER, Bruce (1995), "Lessons from the US Unemployment Insurance Experiments". *Journal of Economic Literature*. 33(1): 91-131.
- MICHALOPOULOS, Charles, Howard BLOOM y Carolyn HILL (2004), "Can Propensity Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?" *Review of Economics and Statistics*, por publicarse.
- MOFFITT, Robert (1991), "Program Evaluation with Nonexperimental Data". *Evaluation Review*. 15(3): 291-314.
- MOFFITT, Robert (2003), "Remarks on the Analysis of Causal Relationships in Population Research". Manuscrito no publicado, Johns Hopkins University.
- NEWKEY, Whitney, James POWELL y James WALKER (1990), "Semiparametric Estimation of Selection Models: Some Empirical Results". *American Economic Review*. 80(2): 324-328.
- NEWHOUSE, Joseph (1994), *Free for All: Lessons from the Rand Health Insurance Experiment*. Cambridge, MA: Harvard University Press.
- NOLL, Roger y Andrew ZIMBALIST (1997), *The Economic Impact of Sports Teams and Facilities*. Washington, DC: Brookings Institution.
- ORR, Larry (1998), *Social Experiments: Evaluating Public Programs with Experimental Methods*. PLACE: Sage Publications.
- OSBORNE, David y Ted GAEBLER (1992), *Reinventing Government: How The Entrepreneurial Spirit is Transforming the Public Sector*. Boulder, CO: Perseus.

- PENCE, Karen (2003), "Foreclosing on Opportunity: State Laws and Mortgage Credit". FEDS Working Paper 2003-16.
- PORTNEY, Paul (1994), "The Contingent Valuation Debate: Why Economists Should Care". *Journal of Economic Perspectives*. 8(4): 3-17.
- PUHANI, Patrick (2000). "The Heckman Correction for Sample Selection and Its Critique". *Journal of Economic Surveys*. 14(1): 53-68.
- RAVALLION, Martin (2001), "The Mystery of the Vanishing Benefits: An Introduction to Evaluation". *World Bank Economic Review*. 15(1): 115-140.
- ROSENBAUM, Paul y Donald RUBIN (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects". *Biometrika*. 70(1): 41-55.
- ROSENBAUM, Paul y Donald RUBIN (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score". *American Statistician*. 39: 33-38.
- SIMS, Christopher (1996), "Macroeconomics and Methodology". *Journal of Economic Perspectives*. 10(1): 105-120.
- SMITH, Jeffrey (2000), "A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies". *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 136(3): 247-268.
- SMITH, Jeffrey y Arthur SWEETMAN (2001), "Improving the Evaluation of Employment and Training Programs in Canada". Texto inédito, University of Maryland.
- SMITH, Jeffrey y Petra TODD (2004), "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?". *Journal of Econometrics*, por publicarse.
- SMITH, Jeffrey y Alexander WHALLEY (2004), "Are Participants Good Evaluators? Evidence from the Job Training Partnership Act". Texto inédito, University of Maryland.
- SNOW, Arthur y Ronald WARREN (1996), "The Marginal Welfare Cost of Public Funds: Theory and Estimates". *Journal of Public Economics*. 61(2): 289-305.
- SWEETMAN, Arthur, William WARBURTON, Rob MCPHEE y Rebecca WARBURTON (2003), "Disability Status: Impacts on Health and Welfare Dependence". Manuscrito no publicado, Queen's University.
- US DEPARTMENT OF LABOR (2000), "Core and Customer Satisfaction Performance Measures for the Workforce Investment System". *Training and Employment Guidance Letter*, núm. 7-99. Washington, DC: Employment and Training Administration.
- VAN DER KLAUW, Wilbert (2002), "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Design". *International Economic Review*. 43(4): 1249-1287.
- WALLSTEN, Scott (2000a), "The R&D Boondoggle". *Regulation*. 23(4): 12-17.
- WALLSTEN, Scott (2000b), "The Effects of Government-Industry R&D Programs on Private R&D: The case of the Small Business Innovation Research Program". *Rand Journal of Economics*. 31(1). 82-100.
- WILSON, James (2000), *Bureaucracy: What Government Agencies Do and Why They Do It*. Nueva York: Basic Books.

- WINSHIP, Christopher y Stephen MORGAN (1999), "The Estimation of Causal Effects from Observational Data". *Annual Review of Sociology*. 25: 659-706.
- WOOLDRIDGE, Jeffrey (2001), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- WOOLDRIDGE, Jeffrey (2002), *Introductory Econometrics: A Modern Approach*, 2a. edición. Mason, Ohio: South-Western College Publishing.
- WALLSTEN, Scott (2000b), "The Effects of Government-Industry R&D Programs on Private R&D: The case of the Small Business Innovation Research Program". *Rand Journal of Economics*. 31(1). 82-100.
- WILSON, James (2000), *Bureaucracy: What Government Agencies Do and Why They Do It*. Nueva York: Basic Books.
- WINSHIP, Christopher y Stephen MORGAN (1999), "The Estimation of Causal Effects from Observational Data". *Annual Review of Sociology*. 25: 659-706.
- WOOLDRIDGE, Jeffrey (2001), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- WOOLDRIDGE, Jeffrey (2002), *Introductory Econometrics: A Modern Approach*, 2a edición. Mason, Ohio: South-Western College Publishing.

Capítulo 13

La evaluación de programas del tercer sector

Por
Andrea Westall,
Directora Adjunta, New Economics Foundation,
Londres, Reino Unido

Introducción: implicaciones de las políticas del tercer sector

El tercer sector es crucial dentro de las estrategias de desarrollo económico local. Las actividades de sus organizaciones no sólo tienen diversos impactos positivos de regeneración, sino que se están volviendo más importantes para las organizaciones de desarrollo económico y para el suministro de servicios públicos a nivel local. Las organizaciones del tercer sector contienden a favor de ciertos impactos de “valor agregado”, tales como el capital social y la participación comunitaria, que pueden ser elementos vitales en el desarrollo económico. Se les acreditan además algunas innovaciones que han influido en las nuevas políticas y modelos de desarrollo económico. El modelo de mercado laboral intermedio, por ejemplo, fue una respuesta del tercer sector al fenómeno del desempleo prolongado.

Por lo tanto, se hace necesario evaluar los impactos y resultados de las actividades del tercer sector para poder determinar de qué manera coadyuvan en la regeneración, cuál es la medida o el grado de sus impactos, y cómo podemos apoyar y desarrollar sus particulares “valores agregados”. Los resultados de una buena evaluación pueden ser utilizados para respaldar el desarrollo e impulsar mejoras en el sector, replicar proyectos, refinar o modificar políticas públicas y apoyar el financiamiento adecuado y efectivo de sus actividades.

Sin embargo, hoy día es muy limitada la evaluación de los impactos del tercer sector en el desarrollo económico local. Lo anterior se debe a diversos factores que surgen de limitaciones internas; a los requerimientos de distintos patrocinadores que pueden estar en conflicto, limitar o crear sobrecargas; y a la falta de procesos y sistemas de medición apropiados que interpreten los resultados, a menudo intangibles, de las organizaciones del tercer sector. La evaluación también se torna difícil porque es común que las organizaciones del tercer sector tengan una gran gama de metas y actividades. Generan impactos que son complejos de desenredar de su contexto y de las acciones de otros actores en su mismo sector y en los sectores público y privado.

El presente trabajo resume distintos enfoques de evaluación del tercer sector, destacando la necesidad de desarrollar sistemas de medición que partan de los objetivos y necesidades de una organización y no simplemente de los requerimientos de los donadores y funcionarios públicos. Esto redundará en

evaluaciones más completas y a su vez proporcionará información útil para el desarrollo de políticas.

Dichos enfoques comprenden evaluar el proceso (evaluación formativa) y los impactos ex-post (evaluación acumulativa). Es necesario apuntalar ambos tipos de evaluación para así poder recolectar datos sistemáticamente y crear métodos para la captación de impactos difíciles de cuantificar. También es esencial reconocer la importancia de incorporar a la comunidad y a los beneficiarios en el proceso de evaluación, con el objeto de crear y mejorar las mediciones de resultados, identificar las consecuencias inesperadas, y comprender mejor las causas y efectos. La evaluación participativa también puede crear programas más robustos y efectivos porque contribuye al proceso de regeneración fomentando la participación social de la comunidad local e involucrando a los beneficiarios en la búsqueda de resultados y objetivos.

Se enumeran a continuación varias implicaciones en el diseño de políticas:

- **Una mejor comprensión de cómo las organizaciones del tercer sector contribuyen al desarrollo económico local, puede conducir a cambios de políticas, mejores convenios de asociación y apoyo directo a las organizaciones cuyo impacto es significativo dentro de los objetivos clave del desarrollo.**
- **Se deben combinar metodologías para entender “cómo funciona”, “si funciona bien o mal” y “cómo mejorar”.**
- **Al donar fondos a proyectos de organizaciones del tercer sector, se debe poner más atención a los resultados que a los productos (*outcome versus output*).**
- **Las organizaciones del tercer sector requieren recursos y capacidades para poder evaluar sus actividades adecuadamente. Esto podría implicar destinar recursos financieros a la evaluación o bien hacer recomendaciones sobre los métodos más efectivos, involucrando a las agencias de apoyo.**
- **Las evaluaciones que logran señalar los impactos sociales y ambientales y los relacionan con las metas y el gasto público podrían fomentar formas innovadoras de financiamiento, basadas en la inversión social o en programas de incentivos.**
- **Si las evaluaciones logran identificar beneficios sociales significativos, pero de bajo rendimiento económico, se crean argumentos a favor de la necesidad de subsidiar proyectos cuyos beneficios sociales contribuyan significativamente a las metas de desarrollo local.**
- **Es importante reconocer que las organizaciones del tercer sector tienen ante sí un abanico de requerimientos de evaluación, por lo que resulta más beneficioso apoyar la creación de sistemas internos de recolección de datos que puedan utilizarse en evaluaciones internas y externas.**

- **Comprender los impactos del tercer sector fomentará la “creación de mercados”** por parte del sector público, apoyando así el desarrollo de organizaciones que trabajen conjuntamente a favor de prioridades múltiples.
- **Una buena evaluación puede llevar al sector público a reconsiderar sus estrategias de compras y contratación**, con el fin de aprovechar los múltiples beneficios de los proveedores del tercer sector, como por ejemplo prestar un servicio básico que también reduce el desempleo. Esto significa que los contratistas del gobierno podrían licitar y diseñar contratos que incorporen los valores agregados que el tercer sector aporta a las políticas públicas, y destinar más recursos al tercer sector para apoyarlo en sus actividades de captación de fondos y capacidad de suministro.¹
- **Una mejor comprensión de las organizaciones del tercer sector y de las estrategias de base comunitaria nos permite identificar y reconocer el valor de los enfoques menos dependientes de programas verticales.** Los modelos que involucran a la comunidad y que toman en cuenta las contribuciones del tercer sector y de los sectores privado y público deben ser evaluados adecuadamente para determinar sus impactos relativos dentro del desarrollo económico.

El último punto señala un aspecto importante que no es tratado en este documento. Resulta de crítica importancia reconocer la multiplicidad de actores y de impactos en las asociaciones de desarrollo y en las áreas en desventaja, y medir la sinergia entre estos actores y su relación con los resultados generales. Se debe trabajar más en el desarrollo de un enfoque a la evaluación que permita apreciar los impactos generales de una variada gama de actividades y organizaciones, pero también identificar las contribuciones exclusivas de cada actor individual y de sus interacciones.

¿Qué es el tercer sector?

El concepto de “tercer sector” es controvertido. Por ejemplo, puede equipararse al concepto de “economía social” que tiene la Unión Europea, el cual aglutina a las cooperativas, sociedades mutualistas, asociaciones y fundaciones (o CMAF, por sus siglas en inglés). Otros piensan que esta clasificación es limitada, ya que no toma en cuenta la existencia de organizaciones no lucrativas o empresas sociales que no necesariamente entran dentro de esas categorías.

Efectivamente, el tercer sector describe un universo que incluye términos más familiares —tales como organizaciones no lucrativas, voluntariados, mutualidades, empresas sociales o empresas comunitarias. Estas organizaciones tienden a centrar su atención en objetivos sociales o ambientales en vez de perseguir réditos o ganancias monetarias. Sin embargo, algunas de las empresas

generan utilidades, pero las reinvierten en sus actividades o bien las distribuyen entre los actores o afectados en vez de destinarse accionistas externos o al dueño/administrador.

Su universo es muy diverso y no tiene fronteras claramente definidas, traslapándose con el sector de negocios y con el sector público. Existen, por ejemplo, casos sumamente interesantes de inversiones y organizaciones privadas que además pertenecen parcialmente a la comunidad y al sector público y que a menudo ofrecen formas innovadoras de prestar servicios públicos.²

El tercer sector abarca desde organizaciones que dependen completamente de donaciones hasta las que son cien por ciento autofinanciadas, desde empresas que compiten en los mercados principales hasta las que suministran bienes y servicios no comerciales. Su gama de metas sociales y ambientales es extremadamente diversa, tanto como lo son sus estructuras de gobierno y los actores que participan en cada organización.

Esta diversidad y el hecho de traslaparse con otros sectores implica que es difícil crear marcos evaluativos que difieran conceptualmente de los ya utilizados para otros fines. Sin embargo, puede que dichos marcos requieran algunas modificaciones específicas a las particularidades del tercer sector.

¿Cómo apoya el tercer sector al desarrollo económico local?

La importancia del tercer sector radica en sus aportes a la regeneración en formas que van más allá de la renovación física y la inversión interna. Es ampliamente reconocido que la regeneración y el desarrollo económico son procesos multifacéticos que involucran diversos aspectos: capital social, salud y bienestar social, cultura local y creación de empleos y de nuevos negocios. De hecho, la regeneración beneficia a una población sólo si asegura que los empleos que se generan sean precisamente para sus habitantes, y que la riqueza creada no se “fugue” del área.

El tercer sector pueda desempeñar un importante rol al:

- Incrementar los empleos y las capacidades de los potenciales empleados.
- Apoyar la creación de nuevas empresas, por ejemplo, a través de espacios administrados de trabajo o de organizaciones que respaldan la creación de nuevos negocios.
- Proponer formas incluyentes y novedosas de hacer negocios, por ejemplo, los modelos cooperativos.
- Abordar las necesidades insatisfechas de bienes y servicios, por ejemplo, de vivienda, alimentación y financiamiento.
- Abordar problemas sociales (salud, servicios a refugiados, etcétera).

- **Construir infraestructura a nivel local** —por ejemplo, instalaciones deportivas, teatros o sus similares, soluciones al transporte, etcétera.
- **Crear capital social y un mayor compromiso de la comunidad.**
- **Promover el reciclaje y buscar respuestas para otras necesidades ambientales.**

¿Por qué medir el impacto?

El proceso de medir resultados e impactos es una forma de comprobar si las organizaciones han alcanzado las metas que se propusieron y cumplido los compromisos adquiridos ante sus donantes. También permite a las organizaciones entender “cómo” lo están logrando, lo que a su vez podría constituirse en un apoyo al desarrollo organizativo y a los procesos de obtención de recursos, y ayudar a replicar el modelo en cuestión o bien a cambiar de política. Las innovaciones en el tercer sector muchas veces son una especie de vitrina que refleja las políticas sociales, ambientales y económicas. Una evaluación también puede descubrir los beneficios inesperados de ciertos programas o intervenciones.

Otra justificación de la evaluación es poder rendir cuentas a las partes interesadas y promover una mayor participación de las mismas. Lo anterior es necesario para legitimar el proyecto pero también para crear una sensación de propiedad compartida, permitiendo a los beneficiarios y a otros actores contribuir con sus conocimientos para comprender “cómo” es que están funcionando las cosas y cómo podrían mejorar. Esto significa que los enfoques participativos de evaluación contribuyen a que el proceso de desarrollo económico sea más efectivo, mediante la creación de nuevas maneras de abordar las necesidades locales (generalmente incorporando a los miembros de la comunidad) y de nuevas respuestas ante oportunidades específicas de cambio económico y social a nivel local. Los cambios pueden parecer pequeños y simples, pero tienen el potencial de generar grandes impactos. Por ejemplo, una evaluación sobre el nivel de criminalidad de un área concluyó que la gente creía que los árboles del parque adyacente a sus hogares ayudaba a los ladrones a pasar inadvertidos cuando escalaban las paredes de sus casas. Al mover los árboles de lugar, aumentó la sensación de seguridad (aunque no se registró si los niveles de crimen se redujeron).³

La evaluación también es una parte necesaria del cambio organizativo y de la administración. Apoya al equipo de trabajo y a los directores en el desarrollo y revisión de programas, así como a planear futuras estrategias y formas más efectivas de trabajo.

Por lo tanto, de lo que se habla aquí es de dos tipos de evaluación: uno que enfoca el proceso formativo y que tiene implicaciones en el desarrollo del programa en curso, y otro con enfoque sumativo o acumulativo que demuestra

los impactos finales del programa. No hay ninguna razón por la que estos dos enfoques no puedan combinarse. Frecuentemente, la información recabada mediante la evaluación del proceso suele ser la misma que se necesita para una evaluación acumulativa.

Sin embargo, en el tercer sector hay dirigentes que desdeñan la evaluación. Para algunos, la evaluación es un lujo que sin duda sería de utilidad, si no hubiesen necesidades más apremiantes que se deben satisfacer con muy pocos recursos disponibles.

También se ha argumentado que medir el desempeño podría obligar a las organizaciones a enfocarse demasiado en estar continuamente “mejorando”, cuando de hecho lo que están haciendo ya funciona bien para el tercer sector.⁴

Similitudes y diferencias en los métodos de evaluación del tercer sector

Medir el impacto del tercer sector no es muy diferente a evaluar otros programas o actividades de desarrollo económico local. Es cuestión de emplear las herramientas y técnicas más apropiadas. Sin embargo, la evaluación del tercer sector es a menudo impulsada por la necesidad de mostrar ciertos impactos en particular, de señalar que al hacer las cosas de cierta manera se les otorga un valor agregado, y explicar por qué. Es también importante evaluar globalmente todos los impactos de una organización, no sólo sus actividades específicas. Bien podría haber sinergia entre los proyectos que dirige una organización, o bien su forma de trabajar en sí podría generar valor agregado. Un ejemplo de lo anterior son los proyectos habitacionales en los cuales el inquilino es a su vez dueño. Aparte de prestar un servicio, este tipo de proyecto ha generado comprobadamente otros beneficios para los inquilinos, como confianza en sí mismos y su participación en actividades cívicas.⁵ Mediante el análisis de la misión general de la organización y su efectividad, y no sólo del impacto de proyectos individuales, la evaluación puede también detectar fallas o medir la capacidad general y efectividad de los procesos y de su equipo de trabajo. Los marcos estándar de evaluación de programas probablemente no pueden captar todos los procesos en marcha, ni los intangibles específicos propios de cada organización. Se ha sugerido que los resultados específicos a los que nos referimos deben incluir:

- Cercanía con la comunidad.
- Participación activo de los usuarios y beneficiarios.
- Acceso a y comprensión de los grupos en desventaja.
- Confianza de los usuarios en la calidad de un servicio. Esto puede ser particularmente importante en aquellos servicios en los que es difícil que un usuario o

sus parientes puedan juzgar la calidad; por ejemplo, en el cuidado de ancianos o de niños.

- **Suministro de bienes y servicios no comerciales.**

Estos atributos o actividades generalmente se dan por hecho, o bien son contabilizados en vez de ser demostrados. Tampoco se circunscriben exclusivamente al tercer sector. Por lo tanto, es importante demostrar estos atributos y que las organizaciones puedan medir los impactos y mejorar su desempeño con base en la información que van adquiriendo.

La medición de impactos y el tercer sector

Primero, debemos distinguir la diferencia entre evaluar los impactos de los programas específicos de una organización del tercer sector y evaluar el impacto de la organización como un todo. También está el reto de ordenar los aportes de las diferentes organizaciones en los casos de asociación multipartita.

También debemos saber distinguir entre los procesos empleados por la organización, por ejemplo la co-participación de actores diversos, y sus resultados particulares. Esto es necesario porque puede argumentarse que involucrar a los actores requiere tiempo y además recursos que podrían invertirse mejor directamente. Sin embargo, los procesos pueden mejorar la calidad de los servicios prestados, o bien constituirse en objetivos de gran valor, generando —por ejemplo— un mayor capital social, más participación de la comunidad, y una mayor autoestima entre sus miembros.

Desafortunadamente, las organizaciones del tercer sector a veces no son muy eficientes a la hora de revelar sus impactos o el valor agregado particular que generan. Una de las principales causas es la copiosa cantidad de requisitos de evaluación que imponen los patrocinadores de los programas o los otros financistas, requisitos con diferentes estándares de medición y tiempos de entrega. Lo anterior puede resultar en una sobrecarga administrativa, con costos elevados y con una evaluación parcial de las actividades de una organización. Los requerimientos de evaluación de los patrocinadores se basan sencillamente en productos (*outputs*) y no en resultados y análisis de calidad.

Por lo tanto, muchas organizaciones del tercer sector tienden a percibir la supervisión y la evaluación de los patrocinadores como una carga y no como un proceso útil. Sienten que las evaluaciones externas generalmente tienen poca utilidad como herramientas de aprendizaje y que no transmiten ni comunican la esencia de lo que realmente están haciendo.

Otras barreras que se interponen a la evaluación son:

- **Restricciones de tiempo, de recursos, de destrezas y de conocimiento.**

- La complejidad de muchos enfoques de evaluación.
- La plétora de modelos disponibles conlleva variedad pero también dudas en cuanto a la calidad de las diferentes técnicas de evaluación; hay que saber escoger los modelos más adecuados para cada caso en particular. Adicionalmente, contar con múltiples enfoques de evaluación puede dificultar las comparaciones y referencias cruzadas entre organizaciones similares.
- Ausencia de técnicas apropiadas, particularmente aquellas que puedan evaluar resultados e impactos intangibles. Las empresas sociales, por ejemplo, son organizaciones que operan en los mercados principales pero que emplean principalmente a personas discapacitadas física o mentalmente. Actualmente están estudiando cómo medir sus capacidades para incrementar resultados positivos en el campo de la salud.
- La diversidad de objetivos y resultados puede que necesite una gama de enfoques de evaluación.
- Los distintos requisitos de evaluación según los actores: patrocinadores, beneficiarios y administradores.
- Preocupación de los patrocinadores en el sentido de que su dinero se esté gastando en evaluaciones en vez utilizarse para la actividad a la que se destinó originalmente.
- Creer que la evaluación es un examen o una prueba, en vez de verla como una herramienta que puede aumentar la efectividad.

Además, existen también otros problemas de índole general que se presentan al medir los impactos de regeneración, los cuales incluyen:

- El tiempo que debe transcurrir para que sean perceptibles los resultados e impactos de las intervenciones o iniciativas.
- La abundancia de factores interrelacionados dentro de un contexto particular o entre diferentes grupos de personas, lo cual incide en el éxito o en el fracaso de una iniciativa.
- La dificultad de separar el impacto de otras influencias, locales o generales.

Técnicas de evaluación utilizadas

Las técnicas de evaluación que utilizan las organizaciones del tercer sector van desde las cualitativas hasta las cuantitativas y de las externas hasta las participativas que involucran a los usuarios en la definición de sus propias medidas de evaluación. Entre más participativo es el enfoque, más probable es que la medición en sí se convierta en una parte integral del proceso de regeneración —ya sea ayudar a definir o perfeccionar el programa, o modificar la participación o el grado de compromiso de los actores.

Evaluación participativa. El Seguimiento y Evaluación Participativa (PM y E, por sus siglas en inglés) nació en la década de 1970 en el campo de desarrollo internacional, tras reconocer que la evaluación externa con criterios programados no puede detectar cómo ocurren los cambios ni crear procesos incluyentes que funcionen de acuerdo a los puntos de vista y a las aspiraciones de los más afectados.*

* Para mayor información sobre el PM y E usted puede consultar, entre otras fuentes, un Resumen de Políticas en el *Institute for Development Studies Website: Policy Briefing, Issue 12, November 1998.* www.ids.ac.uk.

Este enfoque incorpora a la población en la selección y diseño de indicadores adecuados para evaluar resultados a nivel local y también en el proceso de recolección de datos. Por lo tanto, contribuye al proceso de renovación mediante la identificación de criterios locales importantes así como el involucramiento de las comunidades en la identificación de soluciones. La evaluación participativa desafía algunas ideas establecidas sobre el análisis y la recolección de datos. En algunos casos, la información recabada puede ser más confiable. Por ejemplo, en *Merthyr Tydfil*, del Reino Unido, la información sobre el crimen que los niños participantes recabaron para un estudio de 1996, resultó ser mejor que la recabada por la policía, posiblemente debido a la tendencia de no mentirle a los niños.⁶

El proceso puede además validarse externamente mediante la selección de indicadores relevantes y muestreos apropiados.

Muchas organizaciones del tercer sector hacen uso de ciertas técnicas de evaluación, pero frecuentemente de manera básica o *ad hoc*. Existe un amplio rango de enfoques de evaluación que pueden ser utilizados pero no es fácil acceder a ellos o no están hechos a la medida específica del caso.

Fortalezas

- Retroalimenta el proceso de regeneración.
- Involucra a las partes interesadas, resultando en una mayor sensación de pertenencia y de compromiso en el proceso que puede incrementar la confianza, la conciencia social y la capacidad de la comunidad.
- Incrementa los conocimientos locales y la experiencia sobre lo que funciona y cómo funciona, aumentando su comprensión de las causas, y por lo tanto el aprendizaje.
- Puede generar información de mejor calidad.
- Involucra a personas que usualmente quedan al margen en las encuestas formales.
- Estimula la acción y el compromiso cívico.

Debilidades

- Requiere tiempo y recursos.
- Problemas relativos a la comparación y solidez.
- Se debe equilibrar la necesidad de indicadores particulares pero a su vez comparables a los de otras áreas, para así poder valorar impactos relativos con indicadores locales específicos a su contexto. Lo anterior es una herramienta invaluable para comprender cómo suceden las cosas y cómo pueden mejorarse. Un ejemplo de una pregunta comparable sería: ¿Se siente usted seguro en su zona? Y una pregunta más específica sería: ¿Se siente usted seguro cuando camina por el parque X? La primera pregunta es general mientras que la segunda es específica, refiriéndose a un sitio específico en donde de ser factible podría llevarse a cabo una intervención.
- La participación no es un hecho, requiere creatividad.
- La participación podría verse sesgada por grupos de presión y residentes más “poderosos” o vociferantes. En términos muy generales, la evaluación subjetiva puede tornarse problemática porque puede cambiar según el momento o el particular contexto en el que se realiza.

Se han escrito varios documentos sobre el tema de la evaluación en el tercer sector. Está, por ejemplo, el Proyecto de Medición del Sector Independiente —un estudio de técnicas de medición de resultados que se emplea en los EU.⁷ Está también el estudio canadiense sobre recursos de evaluación de Bozzo y Hall.⁸

Muchos trabajos se han enfocado al tercer sector y la regeneración, desarrollando un abanico de enfoques y marcos evaluativos que se concentran en medir el valor agregado particular de las organizaciones del tercer sector, en aspectos como la participación comunitaria o el capital social.

Kendall y Knapp (1999) presentaron un enfoque para medir el desempeño que toma en cuenta el valor agregado del sector voluntario —elección, participación, promoción e innovación— así como los criterios normales para evaluar al sector público: economía, eficacia, eficiencia y equidad.⁹ Si bien este enfoque era teórico, fue utilizado en Irlanda del Norte para seleccionar los indicadores adecuados y desarrollar un sistema que midiese el capital social.¹⁰

Otro ejemplo de marcos evaluativos para el tercer sector es el que desarrolló el Scottish Community Development Centre, titulado “Cómo alcanzar un mejor desarrollo comunitario” (ABCD, por sus siglas en inglés).¹¹ Este modelo ve la evaluación como una parte integral del desarrollo de la comunidad, y a la participación comunitaria como elemento vital de la evaluación. Desarrolla una pirámide de resultados que van desde la potenciación y participación personal hasta la creación de comunidades sostenibles.

La New Economics Foundation también ha desarrollado un modelo para medir el capital social, llamado *Prove It!*, que permite el uso de mediciones específicamente locales así como mediciones que pueden compararse en otras áreas y proyectos. Por lo tanto, su información resulta útil en evaluaciones tanto formativas como acumulativas.

Prove It!

Prove It! es un método de evaluación que desarrollaron conjuntamente el Barclays Bank y la entidad Groundwork, una organización ambiental no lucrativa en el Reino Unido, para poder evaluar los cambios del “capital social” como resultado de proyectos de regeneración en áreas específicas. Si bien el concepto de capital social es controversial, este proyecto fue específicamente diseñado para descubrir si los programas para el mejoramiento de las condiciones físicas a nivel local, generaban a su vez otros cambios en la comunidad, por ejemplo, un aumento de las relaciones interpersonales y de los niveles de confianza entre sus miembros. El modelo es accesible a nivel individual y del sector público, y emplea un lenguaje de medición y de aprendizaje, más que de evaluación.

Prove it! mide los efectos, a menudo invisibles, que producen las actividades de regeneración en los habitantes que trabajan activamente en la selección de indicadores y recolección de datos. Algunos indicadores ya existen y otros son de reciente creación para captar cambios de actitud y de actividades (por ejemplo, los niveles de *graffiti*). Se emplean técnicas cualitativas y cuantitativas, incluyendo encuestas y la información local que ya existe. El proceso completo y sus productos están contenidos en un manual muy sencillo de usar.¹²

Algunos de los impactos no intencionales de los proyectos evaluados fueron: una mayor participación comunitaria, ideas para nuevos proyectos, compromiso con las actividades de regeneración y, por supuesto, el disfrute como resultado de participar y aprender del proyecto. También produjeron una mejor comprensión generacional debido a la participación de los jóvenes en la tareas de recolección de datos.

El sistema *Prove It!* se está perfeccionando para hacerlo más riguroso y para que sea un referencial en otras actividades y áreas, reteniendo sin embargo su metodología participativa.

El uso de indicadores

El elemento clave de todos los enfoques mencionados hasta aquí es precisamente la evaluación de procesos y resultados que son difíciles de medir, como por ejemplo, sentir más confianza y menos miedo al crimen. Para no tratar de re-inventar la rueda, podemos recurrir a la comparación y emplear algunos

de los indicadores más comprobados y robustos, los cuales resultan muy útiles. Algunos ya fueron desarrollados, otros están en proceso de desarrollo, y otros aún deberán ser diseñados.

Un buen ejemplo de un indicador innovador y útil es el de la “distancia recorrida”, utilizado por muchas organizaciones del mercado laboral y en proyectos de generación de empleo.¹³ Este concepto se refiere al progreso que va teniendo un beneficiario en su aptitud o capacidad laboral; se trata de resultados que son más fáciles de medir y que son por supuesto atribuibles a la intervención del proyecto. Mide resultados que son más sutiles, tales como la habilidad de relacionarnos con nuestros semejantes, nuestra habilidad organizativa y de administración del tiempo, entre otros, en vez de medir simplemente si alguien encontró o no un empleo. Para las personas que han estado desempleadas por mucho tiempo, el hecho de finalmente conseguir un empleo es un paso gigantesco. Se necesita tiempo y esfuerzos para que esta persona se convierta en buen sujeto de empleo, por decirlo de alguna manera. Aumentar su capacidad de empleo involucra, de hecho, un rango de destrezas y cambios de actitud, y la distancia que deberán recorrer antes de lograr un empleo será mayor para unos que para otros.

Tales mediciones son por lo tanto más adecuadas como indicadores del éxito de una intervención, versus medir sencillamente la obtención o no obtención de empleo. También resultan vitales para asegurar que los donadores reconozcan el valor agregado real de los proyectos, y que no tracen metas que podrían sesgar el comportamiento de las organizaciones, de tal manera que al final éstas tengan que trabajar sólo en beneficio de las personas que son buenos sujetos de empleo para poder cumplir con las metas de sus patrocinadores.

Los resultados esenciales medidos se relacionan con habilidades clave de trabajo, como la comunicación; habilidades de actitud tales como una mayor motivación y confianza; habilidades personales, tales como la puntualidad; habilidades prácticas tales como saber llenar formularios. Pueden darse entonces resultados específicos a grupos objetivo distintos, por ejemplo, aquellos conformados por personas con deficiencias de aprendizaje.

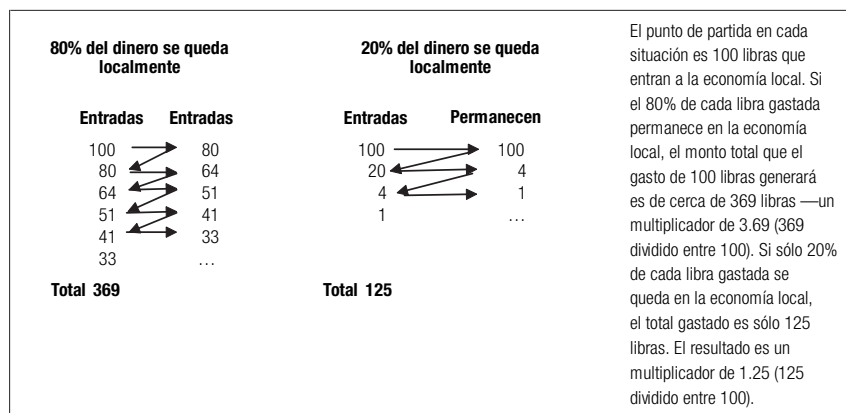
Un estudio del Departamento de Educación y Empleo (DfEE, por sus siglas en inglés) del Reino Unido determinó que si bien los enfoques para medir la “distancia recorrida” son útiles, también hay otras técnicas en uso. Esto torna difícil la comparación.¹⁴ Además se tiene el problema de la subjetividad en asesores y beneficiarios, y el de atribuir los cambios exclusivamente a un proyecto. El resultado del estudio del DfEE fue la preparación de mejores guías para la medición, y reconocer la necesidad de enfoques focalizados. Este ejemplo claramente ilustra las posibilidades y los retos que se presentan inevitablemente en la creación de indicadores.

La New Economics Foundation presentó un resumen de los diversos indicadores potenciales y de sus aplicaciones en una publicación sobre indicadores comunitarios y en el manual *Prove it!* anteriormente referido. Hay también un sitio disponible en internet que describe 450 indicadores que ya han sido utilizados y comprobados en el mundo entero.¹⁵ La referida fundación también desarrolló recientemente un indicador específico, el Multiplicador Local 3, para que las empresas y organizaciones del tercer sector y los gobiernos puedan evaluar el impacto de sus actividades económicas y su contribución a la economía local.

Multiplicador Local 3 (LM3, por sus siglas en inglés)

El siguiente ejemplo ilustra el multiplicador LM3 (véase figura 13.1): Supongamos que 100 libras esterlinas ingresan a la economía local. Si el 80 por ciento de cada £1 se queda en la economía local, el monto total de gasto generado por esas £100, tras cinco rondas consecutivas, será de £369. En este caso el multiplicador será 3.69. Dicha cifra se obtiene mediante la división del monto total de gasto entre el ingreso inicial. No obstante, si sólo el 25 por ciento de cada libra esterlina se queda en la economía local, el total gastado después de cinco rondas será de sólo £125. El multiplicador obtenido será de 1.25 (125 dividido entre 100).

Figura 13.1 **Ejemplo de un multiplicador local**



Para simplificar, el LM3 solo utiliza tres rondas de gasto. Sin embargo, podemos emplear más rondas, y al extrapolar las tasas de gasto se generarán medidas más precisas.

En la práctica, primero se analizan los ingresos de la organización o de la actividad. Se le da seguimiento a este dinero utilizando estados de cuentas o

cuestionarios para determinar cómo es gastado localmente y cuánto es gastado afuera. Se sigue entonces al siguiente nivel de gasto y se observa cómo es gastado por los empleados o los proveedores que lo recibieron. La ronda 1 son los ingresos; la ronda 2 es el gasto; y la ronda 3 es el gasto del personal y los proveedores. Por último se calcula el LM3 sumando las tres rondas y dividiendo el total de la suma por el ingreso original.

A continuación se ilustra este proceso con dos ejemplos del tercer sector:

Eden Community Outdoors es una empresa social que brinda oportunidades a la gente joven, por ejemplo, a través del establecimiento de nuevos negocios o de proyectos ambientales. Dicha empresa calculó que su multiplicador local es de 2.02, o sea que por cada £10 que reciben, la compañía gana 10.20 libras adicionales.

Ibstock Community Enterprises es una organización de regeneración con múltiples actividades en una comunidad rural. Sus miembros deseaban medir el impacto de un cajero automático que habían instalado en respuesta al cierre de la sucursal bancaria local. Descubrieron que los habitantes de la comunidad gastaban entre el 50% y el 67% de su dinero localmente. Lo anterior ilustra la importancia de buscar intervenciones que lleven dinero a la economía y que ayuden a que este dinero sea invertido en bienes y servicios locales.

Esta herramienta es de particular interés para las organizaciones del tercer sector porque les permite demostrar el impacto económico de sus actividades y también encontrar formas prácticas de elevar su “multiplicador”. Una organización puede identificar formas para que el dinero que gastan siga canalizándose hacia las actividades locales y por lo tanto contribuir a la generación de empleos y riqueza. Esto es particularmente importante en áreas en desventaja donde frecuentemente es muy difícil determinar si el dinero de la regeneración se está quedando en la comunidad o si se está fugando en forma de salarios o en pagos a proveedores externos.

El LM3 no es exacto, pero es suficientemente sencillo para que las organizaciones y comunidades lo usen para evaluar sus impactos. También puede usarse para analizar los beneficios de la inversión interna o los impactos de las compras y adquisiciones del gobierno o las empresas locales. Otro ejemplo es el de beneficios o prestaciones sociales. Cuando el *Newham Council* en Londres ensayó esta técnica, el multiplicador fue de 1.77. En otras palabras, por cada £1 de gasto social ejecutado, en la comunidad se generó £1.77. La implicación es que las estrategias para incrementar los beneficios de asistencia social podrían ayudar a la creación de riqueza y de empleos en áreas en desventaja.

Desde un punto de vista más general, el LM3 puede usarse para demostrar cómo pueden modificarse las políticas y las prácticas, particularmente las de regeneración y desarrollo económico local, para promover una mayor circulación del dinero o del financiamiento.

Actualmente, la técnica sigue desarrollándose. Los diseñadores de políticas y profesionales en el Reino Unido discuten sus implicaciones. Es una herramienta potencialmente útil en localidades en desventaja o en áreas rurales donde el mantenimiento de los servicios locales adquiere a menudo una importancia crítica.

Sistemas organizativos de evaluación

Existe a su vez una gama de sistemas orientados hacia los resultados organizativos y la participación de los actores, que van más allá de la evaluación de procesos. Este tipo de enfoque, sin embargo, necesita una gama de indicadores y medidas como las anteriormente referidas que permitan efectivamente la creación de sistemas para recolectar y analizar información de diversas maneras.

Dos de estos enfoques se describen a continuación: el primero es un sistema de auditoría social y el segundo es un sistema de administración de información desarrollado por OASIS, del Roberts Enterprise Development Fund en Estados Unidos.

Contabilidad social y auditoría social

La contabilidad social y la auditoría social se remontan a los inicios de la década de 1970. Fueron perfeccionadas en los años noventa por *Traidcraft* y la *New Economics Foundation*, para convertirse en un modelo voluntario que nace del compromiso de las partes interesadas. Hoy día existe ya el Instituto de Responsabilidad Social y Ética (ISEA, por sus siglas en inglés), un cuerpo internacional que promueve mejores prácticas y estándares que han sido utilizadas por grandes corporaciones como la Shell. Este enfoque también se promueve en diferentes programas internacionales financiados por la Unión Europea.

La contabilidad social es un marco metodológico para diferentes tipos de evaluación de impactos. En términos sencillos, el proceso es tal que primero identifica los objetivos sociales y valores de la organización que serán el punto de referencia para valorar cualquier actividad. Luego se define quiénes son los actores participantes y se establecen indicadores de desempeño social, económico y ambiental. Algunos de estos indicadores son diseñados por los actores mismos. El proceso de contabilidad involucra la recolección de datos subjetivos y objetivos que luego pasan a una auditoría social, cuyos resultados son verificados por un asesor independiente y finalmente son publicados.

Los progresos en el campo de la contabilidad social han permitido que las organizaciones pequeñas del tercer sector puedan usar el modelo. En Escocia, un enfoque de aplicaciones variadas fue desarrollado para organizaciones comunitarias; y la entidad Social Enterprise Partnership desarrolló el primer programa europeo de auditoría social que incluye un manual que ya fue traducido a varios idiomas.¹⁶

La contabilidad social combina enfoques cualitativos y cuantitativos. Puede mejorar la rendición de cuentas, otorgar poderes a los actores, permitir la planificación estratégica y demostrar impactos. Crea un proceso mediante el cual se puede involucrar a todos los grupos importantes en la identificación de necesidades y soluciones.

No obstante, uno de los problemas de este enfoque es que requiere de muchos recursos, y se necesitan productos más accesibles. La New Economics Foundation, en coalición con otras organizaciones, actualmente está trabajando en una nueva herramienta de auditoría social llamada Ethical Explorer, que será muy fácil de usar en internet, con el soporte adecuado. Este sistema en línea suministra proformas sencillas, opciones para indicadores y herramientas analíticas para reducir el tiempo y esfuerzo administrativo que conlleva una auditoría social y la producción de un informe. Un prototipo ya está disponible para poderlo examinar y comentar.¹⁷

En Estados Unidos, la entidad Roberts Enterprise Development Fund ha trabajado en la creación de sistemas para evaluar los impactos del tercer sector. Ha centrado su atención en modelos que se relacionan con la generación de empleos, pero sus técnicas tienen otras aplicaciones.

Valoración actualizada de impactos sociales, OASIS (por sus siglas en inglés)

La entidad Roberts Enterprise Development Fund ha lanzado un nuevo modelo para el capital social en riesgo de organizaciones no lucrativas, que les permite desarrollar su capacidad de medir y usar la información sobre los impactos sociales de sus programas, y así evaluar y valorarlos eficazmente. Desean perfeccionar este modelo y que pueda convertirse en un modelo de evaluación estándar en las organizaciones no lucrativas.

El modelo OASIS permite que las organizaciones den seguimiento a los servicios individuales que prestan y a los impactos sociales de sus programas.¹⁸ Las organizaciones no lucrativas tienen por lo tanto la posibilidad de evaluar con más precisión sus servicios, recursos e impactos, lo que les permite calcular la rentabilidad social de su inversión (ver sección *Presentación de los datos*).

En efecto, este método es un sistema informativo de administración social que permite a las organizaciones ser más eficientes y recolectar información en forma sistemática para sus donadores y evaluadores.

Los procedimientos no son sencillos. El método de OASIS es intenso en infraestructura y tecnología e implica tener que re-estructurar significativamente una organización.

Rubicon

Rubicon es una organización en California que ayuda a personas con desventajas, proporcionándoles vivienda accesible, empleo, capacitación laboral, servicios de salud mental y otros apoyos. Sus programas incluyen servicios de apoyo y empresas con vocación social que emplean a adultos de bajos ingresos y con limitaciones físicas, contando con una panadería y una empresa de servicios domésticos.

Rubicon gastaba mucho tiempo y dinero preparando informes y tenía cerca de ochenta donadores. Sin embargo, estas actividades de evaluación no les daban una perspectiva amplia, ni a ellos mismos ni a sus donadores. También tenían el problema que muchos de sus servicios eran integrados y se necesitaba un sistema de administración de casos que fuese flexible y que pudiese dar seguimiento a la gente y disminuyese la recolección de datos. Desarrollar un sistema para recabar, ingresar y reportar información —llamado REDF's CICERO por sus siglas en inglés— les llevó tres años. Sin embargo, se registraron efectos positivos. El sistema los llevó a cambios internos, incrementando su capacidad de evaluar qué tan idóneos eran los programas, y mejorando los servicios y la forma de integrarlos. Actualmente, todos sus contratos e informes para los donadores se basan en estos datos y es más fácil reconocer los indicadores clave del éxito del programa.

Presentación de los datos

Los resultados de una evaluación pueden presentarse en una variedad de formas y a una variedad de audiencias. La auditoría social es una manera exhaustiva de resumir la información y de presentarla en un formato accesible a los interesados. Los elementos de la información recabada pueden servir para desarrollar una serie de actividades de *benchmarking* (evaluaciones comparativas con otros programas) y rigurosos diseños experimentales de mayor duración.

Un enfoque interesante que está siendo desarrollado actualmente en EU y el Reino Unido es el del rentabilidad social de la inversión.

Rentabilidad social de la inversión

Este enfoque permite a una organización demostrar su rentabilidad económica y su rentabilidad social, y crear una rentabilidad “mixta” que puede usarse para comparar proyectos o inversiones según sus respectivos impactos y resultados.

La técnica fue desarrollada en los EU por la entidad Roberts Enterprise Development Fund (REDF). En efecto, la rentabilidad social sobre la inversión (RSI, o SROI, por sus siglas en inglés) le asigna un valor monetario a los impactos sociales creados. Un ejemplo son los beneficios sociales de los proyectos que generan empleos para personas con discapacidad. Los beneficios sociales considerados aquí no se refieren solamente al empleo en sí (y resultados como un mejor estándar de vida, auto confianza, independencia, etcétera) sino que también a mejorías en la salud como consecuencia de tener empleo. La rentabilidad económica en una organización es frecuentemente medido como una rentabilidad sobre la inversión. La rentabilidad social es el valor de los efectos sociales, ambientales y económicos creados por la organización. En algunos casos, puede llegarse a un aproximado de estos impactos mediante la estimación de cambios futuros en el gasto público, tales como el grado de incremento en la recaudación tributaria o la reducción del pago de prestaciones.

Más específicamente, el enfoque de la rentabilidad social de la inversión pretende medir cuál es el valor actual neto del flujo futuro de costos y beneficios, y compararlo con la inversión inicial. El fondo REDF presentó un trabajo que explica detalladamente cómo se debe hacer este cálculo.¹⁹ A manera de ejemplo del enfoque SROI en la práctica, a continuación se presenta el caso de Pedal Revolution, de San Francisco. Esta organización entrena a jóvenes en la venta de bicicletas y reparaciones. El proceso SROI requiere la medición de las rentabilidades económicas y sociales generadas, y debe crear un índice que compare el monto de la inversión con el valor creado. Un índice 1 significa que por cada dólar (euro, etcétera) invertido, se ha generado un dólar (EURO, etcétera) de valor. Si el índice es mayor que uno, eso significa que la inversión ha generado un valor mayor al de la inversión. En el caso de la Pedal Revolution, el índice de la rentabilidad económica es de 9.6. Si se incluye todo el valor adicional que generan sus actividades sociales, el “índice mixto” de rendimiento llega a un 32.5, lo cual ilustra el inmenso valor que tienen los impactos sociales de la organización.

El fondo REDF reconoce que este enfoque se basa en pasados análisis de costo beneficio. Pero intenta seguir avanzando, invitando a profesionales y expertos a desarmar el modelo, por así decirlo, con el fin de poder perfeccionarlo.

El modelo SROI debe identificar y medir una amplia gama de resultados, tanto firmes como blandos. El modelo depende también de la utilización de evaluaciones participativas cuyo efecto sea generar información para poder determinar lo que se necesita medir, particularmente las consecuencias no intencionales de ciertos programas, lo cual podría generar importantes ahorros.

La New Economics Foundation se ha referido a al modelo SROI como un enfoque “estrecho”. Actualmente, se están buscando maneras de “ampliar” el modelo SROI, lo cual permitirá que las organizaciones cuenten con un sistema

no-monetario de relacionar los resultados sociales y financieros para poder captar todos los impactos y no solamente aquellos que fácilmente se atribuyen a cambios en el gasto público. Ejemplos de tales mediciones vienen a ser, por mencionar algunos, la satisfacción del usuario, el capital social o la confianza.

El modelo SROI no sólo permite evaluar plenamente la optimización de los recursos. También podemos usarlo para mostrar la importancia y los impactos de organizaciones que generan poco valor económico, pero cuyo valor social es extremadamente alto. Esto es particularmente importante para aquellas actividades que pueden requerir algún tipo de subsidio porque no son autofinanciables en un 100%. El análisis resulta provechoso para el sector público y otros donantes, pues proporciona parámetros para poder determinar a cuánto debe ascender un subsidio para que pueda alcanzar cierto nivel de rentabilidad social. En algunos casos el subsidio podría en efecto ser una inversión porque, a largo plazo, sobrevendrían ganancias monetarias.

El modelo SROI también es importante porque permite a los administradores seguir la trayectoria de los proyectos y respaldar las aplicaciones de los recursos financieros. Además permite a los diseñadores de políticas tener una clara perspectiva del impacto de las donaciones.

El modelo se sigue desarrollando porque aún tiene varios inconvenientes. Por ejemplo, resulta difícil comparar proyectos aparentemente similares, debido a que éstos utilizan enfoques levemente distintos dirigidos a diferentes grupos objetivo. Actualmente se está experimentando ponderar los diferentes rendimientos. Por ejemplo, los impactos de ayudar a conseguir un empleo a personas que han estado desempleadas durante seis meses, podrían medir menos que si se tratase de desempleados de dos años. Sin embargo, tales enfoques podrían tornarse bastante subjetivos. A menudo también es difícil identificar y cuantificar correctamente los ahorros en el gasto público. Debe tenerse el mismo cuidado con los índices bajos, porque podrían significar que una organización de hecho está generando grandes impactos sociales que son difíciles o imposibles de cuantificar en términos de dinero. También hay problemas al atribuir causas y efectos, debiéndose evitar la doble atribución. Actualmente la New Economics Foundation está trabajando en estos temas para perfeccionar el modelo.²⁰

Este tipo de análisis puede ser útil para los modelos que se basan más en la inversión que en el otorgamiento de los fondos y el apoyo a los programas en sí. Los donantes están gastando por el valor creado, y no simplemente compensando sus gastos. El modelo SROI promueve el financiamiento con base en resultados, no con base en productos. También tiene implicaciones en la toma de decisiones, internamente y de los donantes, y puede generar nuevos sistemas de financiamiento social o de gasto social, en los que los recursos financieros viren de las donaciones de corto plazo hacia programas con incentivos por resultados.

Conclusiones

Este documento ha destacado los retos y oportunidades en la evaluación de impactos del tercer sector para el desarrollo económico local. Los responsables de formular políticas deben comprender mejor estos impactos, y apoyar más eficientemente al tercer sector en la medición de resultados. Si bien es fácil evaluar los programas verticales o crear mediciones para evaluar el gasto público, es evidente que no son las únicas soluciones para el desarrollo económico local, particularmente en las áreas con más desventajas. Es precisamente la gama de diferentes actividades y acciones que se emprenden en un área, incluyendo las del tercer sector, y la sinergia entre todos, lo que traerá cambios a largo plazo. Muchos de los impactos serán intangibles. Evaluar esta desordenada realidad es difícil, pero necesario, para asegurar que el dinero público sea utilizado eficazmente y no que, como este informe ha insinuado, los fondos se fuguen fácilmente.

Notas

1. Esto no significa argumentar necesariamente a favor de predeterminadas adquisiciones, pero sí a favor de incluir un mayor rango de resultados potenciales que cualquier proveedor pueda suministrar.
2. Véase como ejemplo, Westall A. (2001), *Value-Led Market-Driven: Social enterprise solutions to public policy goals*, Institute for Public Policy Research.
3. Véase Walter P., J. Lewis, S. Lingayah and F. Sommer (2000), *Prove It: Measuring the effects of neighbourhood renewal on local people*, NEF, www.neweconomics.org.
4. Véase, por ejemplo, Wainwright S. (2002), *Measuring Impact: A guide to resources*, National Council for Voluntary Organisations (NCVO).
5. Encontrado en un estudio de Price Waterhouse para el entonces Departamento del Medio Ambiente(1995): *Tenants in Control: An evaluation of tenant-led housing management organizations*.
6. S. Platt y A. Treneman, *The Feel Good Factor: a citizens' handbook for improving your quality of life*, Channel 4 Television, Londres, 1997.
7. E. Morley, E. Vinson y H. Hatry (2001), *Outcome measurement in nonprofit organisations: Current practices and recommendations*, Independent Sector, www.independentsector.org.
8. S.L.Bozzo y M.H. Hall (1999), *A review of evaluation resources for nonprofit organisations*, Canadian Centre for Philanthropy.
9. J. Kendall y M. Knapp (1999), *Measuring the performance of voluntary organization activities*, Belfast Voluntary Activity Unit.
10. M. Morrissey y P. McGinn (Community Evaluation Northern Ireland) (2001), *Evaluating community-based and voluntary activity in Northern Ireland*, The Voluntary Activity Unit, DSD.
11. Este enfoque puede encontrarse en la página de internet del Scottish Community Development Centre, cuyo sitio es www.scdc.org.uk. Achieving Better Community Development (ABCD).

12. Este manual está disponible en el sitio de la New Economics Foundation: www.neweconomics.org o en una copia impresa. P. Walter, J. Lewis, S. Lingayah y F. Sommer (2000), *Prove It: measuring the effects of neighbourhood renewal on local people*. El modelo está siendo ampliado y la información estará disponible en el sitio web.
13. Dewson S. y otros (2000), *Guide to Measuring Soft Outcomes and Distance Travelled*, The Institute for Employment Studies, www.esfnews.org.uk/evaluation/documents/distance.pdf.
14. S. Dewson, J. Eccles, N.D. Tackey and A. Jackson (2000), *Measuring Soft Outcomes and Distance Travelled: A Review of Current Practice*, DfEE Research Report RR219.
15. El sitio web de los indicadores de comunidad se encuentra en www.rprogress.org. La publicación de la New Economics Foundation es *Communities Count! A step by step guide to develop community-based sustainable indicators*. El manual está disponible en el sitio web www.neweconomics.org.
16. Véase para ejemplos www.cbs-network.org.uk y www.socialenterprise.co.uk
17. Véase www.ethicalexplorer.org para mayores detalles. *Ethical Explorer* será lanzado en el otoño de 2003.
18. Véase www.redf.org para detalles de OASIS y el ejemplo de Rubicon.
19. Véase www.redf.org. El documento que explica el método del SROI es el *SROI Methodology Paper*. Hay otras fuentes y artículos relacionados con el SROI en este sitio web.
20. Nuestras conclusiones iniciales sugieren que un análisis de las partes interesadas es de importancia crítica, seguido por un análisis insumo-producto y de resultados. Luego debe crearse un modelo causa-efecto. La exactitud de la información y los costos de recopilación se compensan mutuamente. Pueden utilizarse promedios nacionales en vez de datos actuales, pero por el momento no hay fuentes accesibles para eso.

Capítulo 14

Aspectos metodológicos y prácticos para la evaluación de pactos territoriales. La experiencia de Italia*

por
Paola Casavola,
Departamento para el Desarrollo - Unidad de Evaluación,
Ministerio de Economía y Finanzas
Roma, Italia

* Deseo agradecer a Francesca Utili por nuestra investigación común sobre el tema y a Laura Tagle por las múltiples discusiones sobre la evaluación de los aspectos del desarrollo local. La responsabilidad de las opiniones expresadas en este trabajo recae exclusivamente en el autor.

Pactos territoriales italianos

Un pacto territorial es un instrumento específico de política dirigido a promover el desarrollo local a través de incentivos financieros dirigidos a un conjunto de proyectos locales integrados, los cuales son diseñados por una alianza entre actores locales (públicos y privados). En algunos casos los recursos financieros para la asistencia técnica han estado disponibles para la alianza.¹

En Italia, para septiembre de 2002, el número de pactos territoriales aprobados y considerados elegibles para recibir apoyo financiero es bastante grande. Existen 230 pactos, de los cuales 220 (incluidos 91 especializados en agricultura y pesca) fueron seleccionados a través de un procedimiento a nivel nacional, y los 10 restantes se escogieron durante la actualización en la programación de la Estructura de Apoyo a la Comunidad OB.1 1994-1999 sobre la base del procedimiento acordado con la Comisión Europea. Los dos procedimientos difieren. En el procedimiento nacional la selección de los proyectos de pactos (iniciado por una licitación pública nacional) finaliza simultáneamente con la aprobación del proyecto del pacto general y de todas las iniciativas particulares incluidas. El procedimiento nacional se ha repetido y modificado con el curso de los años, por lo que ahora tenemos un conjunto diferente de pactos territoriales nacionales aprobados. El procedimiento de la Comisión Europea fue implementado en una sola ocasión. Dicho procedimiento fue realizado en dos etapas. Primero, 10 proyectos generales y territorios fueron escogidos, posteriormente, a través de licitaciones territoriales, se seleccionaron las iniciativas individuales.²

Aunque los pactos territoriales individuales no cubren áreas muy grandes, dicho instrumento es tan generalizado que una buena parte del territorio nacional se ve afectada (véase figura 14.1).

Los pactos territoriales en Italia mantienen una reputación muy diversa. Entre los políticos, los observadores en general y los expertos en desarrollo territorial hay quienes son sus feroces enemigos y otros sus decididos defensores. Sin embargo, la mayor parte del debate (y la mayoría de los argumentos a favor o en contra) no se ha dado sobre la base de evidencias científicas, monitoreos rigurosos o estudios de evaluación. El debate descansa principalmente en conjeturas, experiencias directas y, a menudo, en prejuicios. En lo que resta

Figura 14.1 Italia: áreas cubiertas por pactos territoriales* (septiembre 2002)



* Los 91 pactos especializados en agricultura no están incluidos (dado que en su mayoría se traslapan con otros pactos).

del documento no consideramos este debate general de forma exhaustiva, principalmente el realizado por los medios de comunicación. En lugar de esto, nos concentraremos en qué y por qué quisiéramos aprender de las evaluaciones de los pactos territoriales e informar lo que hasta el momento se ha hecho al respecto.

Objetivos de los pactos territoriales y preguntas sobre la evaluación

El pacto territorial está construido sobre la idea de que el financiamiento a una alianza de actores con un proyecto, puede reunir agentes individuales, brindando impulso a las economías de aglomeración y asociaciones competentes. En teoría, si el resultado deseado es promover las economías de aglomeración y la capacidad colectiva, dicha línea de acción debe ir más allá que el financiamiento de proyectos individuales valiosos pero que no están coordinados con los actores privados o con las entidades gubernamentales (tal es el caso de los incentivos comunes para inversiones privadas, el financiamiento de iniciativas públicas únicas como infraestructura, capacitación o servicios comunales y sociales).

La idea no es nueva en cuanto a la promoción del desarrollo territorial local. Dicha idea está basada en las conclusiones teóricas y empíricas que muestran que las áreas donde el desarrollo ha sido estimulado, a menudo mantienen interrelaciones sociales y económicas densas. Estas interrelaciones pueden presentarse como formales, informales, dirigidas por el mercado o institucionales. Las conexiones entre las actividades productivas y los servicios relacionados también se armonizan entre los actores locales relevantes. Las economías locales basadas en aglomeraciones coherentes de una variedad de actividades a menudo han probado estar asociadas con la prosperidad local, comparable o aún más duradera, que a las formas de desarrollo territorial provenientes de la sola presencia de empresas y plantas de gran tamaño. Esto es, por ejemplo, el caso de los citados distritos industriales italianos. Estos han sido estudiados exhaustivamente por economistas y científicos sociales. Sin embargo, los estudios se concentran en la evolución natural y el equilibrio, de hecho sus conclusiones no respaldan necesariamente la idea de que es posible para un funcionario promover o acelerar el desarrollo local a través de la concepción de incentivos para que una alianza se forme o progrese rápidamente. La necesidad de evaluar proviene en parte de esta última consideración.³ La pregunta general de la evaluación siguiendo esta línea de razonamiento es si, hasta dónde y en cuáles circunstancias el funcionario puede inducir o acelerar la dinámica de desarrollo local.

Con el fin de identificar las preguntas relevantes específicas de la evaluación de pactos territoriales, resulta útil considerar los objetivos explícitos de la política. La política tiene dos metas específicas consideradas como mecanismos cruciales para inducir o aumentar el desarrollo local. La primera es respaldar la

puesta en marcha de un proyecto sólido y localmente arraigado, integrado por un conjunto de diversas iniciativas que involucren responsabilidades de parte de los actores públicos y privados. Si el proyecto es exitoso, se espera que éste genere externalidades positivas hacia la economía del área. La segunda meta es promover la formación de una asociación local sólida, un grupo de actores que pudieran, a través del ajuste e implementación del proyecto original, aprender a cómo interactuar entre sí y promover un desarrollo mayor. Este segundo objetivo es clave cuando los cambios son permanentes y están relacionados con la creación de una asociación local duradera.

Estos dos objetivos de los pactos territoriales generan dos conjuntos de preguntas de la evaluación. El primero se relaciona con la naturaleza del financiamiento del proyecto integrado (por ejemplo, ¿bajo qué circunstancias el mecanismo de la política probó ser exitoso en inducir y seleccionar para apoyo financiero, un buen proyecto? ¿El proyecto está bien definido y arraigado en el conocimiento real del potencial de un territorio? ¿Es capaz de impulsar un proceso de desarrollo local? ¿Cómo?, etcétera). Un conjunto de preguntas un tanto diferente está relacionado con la naturaleza de la alianza local que el pacto promoverá (¿Es una buena alianza? ¿Ayuda, induce o impulsa el proceso de las relaciones de confianza entre los actores? ¿Ha aumentado la capacidad colectiva para la resolución de problemas? ¿En qué sentido?, etcétera).

Las preguntas arriba esbozadas están relacionadas con los mecanismos específicos de desarrollo que el pacto debe activar directamente (buenos proyectos y capacidad institucional). Sin embargo, muchos desarrolladores de política están interesados principalmente en los resultados económicos finales de los territorios en los que el pacto fue implementado. En otras palabras, existe un tercer conjunto de preguntas más explícitas como del tipo que se encuentra a continuación: ¿han los pactos promovido el desarrollo local, el crecimiento de las empresas y el aumento en las oportunidades de empleo? De hecho estas últimas preguntas son importantes, pero también pueden ser engañosas si no se tratan correctamente. Dos aspectos deben ser considerados: el tiempo transcurrido entre la implementación del proyecto original y los efectos deseables de las externalidades positivas y el hecho que este tipo de preguntas son en verdad muy generales. En la discusión de la evaluación de los pactos territoriales en Italia este último punto es particularmente importante, dado que los pactos son implementados en territorios (especialmente en el Mezzogiorno italiano) donde ya existen otras políticas, las cuales se apoyan en otros mecanismos. La necesidad de desenredar los efectos provenientes de, o acumulados por, diferentes tipos de intervención es de importancia primordial en la discusión sobre los métodos de evaluación. Por lo tanto debemos mantener esto en mente cuando observemos los pactos territoriales y así dejar claro qué esperamos de los ejercicios de evaluación.

El ambiente cultural e institucional para realizar los ejercicios de evaluación: teoría y práctica

Así como en otros contextos podemos evaluar tanto para el aprendizaje (la forma en que el instrumento trabaja en realidad, dónde funciona bien, dónde no funciona y por qué) como para la rendición de cuentas⁴ (para quienes otorgan los fondos). En el caso italiano, sin embargo, la mayoría de las políticas no son evaluadas *ex post* (o a lo largo del curso de las mismas) y el monitoreo sistemático de la política se ha iniciado sólo recientemente. Hasta cierto punto esto representa una buena oportunidad para los pactos territoriales. De hecho, abre una ventana de oportunidad para promover un enfoque de aprendizaje, el cual es menos intimidante para el desarrollador de política y es particularmente recomendable cuando la política en cuestión no tiene antecedentes de evaluación.

Escoger el momento oportuno es un asunto crucial al evaluar el desempeño del pacto. En otras palabras, dadas las características del instrumento y de sus mecanismos es importante que el evaluador sea paciente y tenga visión de largo plazo. Primero, para valorar los resultados finales (tanto en términos de desarrollo como de la capacidad institucional) es necesario esperar a que los pactos acumulen suficientes antecedentes de implementación. Sin embargo, para estar en la capacidad de evaluar los resultados es necesario recolectar cierto volumen de información mientras los pactos se estén implementando y preparar la información estadística necesaria al nivel de detalle territorial deseado. Estas argumentaciones fueron percibidas como muy abstractas hace sólo un par de años atrás. No obstante, dichos conceptos son ahora bien entendidos por muchos de los desarrolladores de la política.

A pesar de la presencia de cierto grado de impaciencia,⁵ en abril del 2002, la administración ha decidido lanzar un estudio que involucra trabajo de campo de un grupo de pactos que se han estado dando durante varios años (subrayando la necesidad de hacer conciencia de que sólo en este caso un estudio puede brindar algunas respuestas sobre el desempeño). Más aún, un nuevo sistema de monitoreo ha sido instalado, el cual debe ser capaz de proveer información precisa de la historia administrativa de las contribuciones financieras a los pactos y ofrecer otra información que en principio podría ser utilizada conjuntamente con otros datos territoriales. Sin embargo, para implementar una verdadera evaluación es necesario que el desarrollador de política presente un conjunto de preguntas completas y detalladas de evaluación, sobre las que esté auténticamente interesado. Si hay un interés en aprender de la evaluación, dichas preguntas no deben ser sólo expresadas en términos de los resultados socioeconómicos ocurridos en los territorios, sino también en términos del funcionamiento del instrumento en sí mismo (¿la configuración del procedimiento para reunir y seleccionar proyectos funciona? ¿Cuáles son los defectos? ¿Cuáles son los incentivos buenos y los malos?).⁶ En este aspecto es necesario mayor avance.

Aun en el estudio iniciado recientemente, a pesar de un incuestionable interés genuino sobre lo que ha ocurrido en los diferentes contextos de los pactos, no se ha tomado el tiempo en detallar preguntas desafiando el papel de la administración en el diseño e implementación del instrumento.

Otro aspecto clave está relacionado con los métodos de investigación de la evaluación. La metodología es crucial al menos por dos razones. Primero, dado que un pacto territorial es un paquete conformado por diferentes elementos (no es una sola política de intervención bien definida, sino un conjunto completo de instrumentos diferentes que se juntan), para aprender de la evaluación debemos conocer los ingredientes clave (qué aspectos de la política son más importantes/ efectivos) y los métodos utilizados deben tener el alcance apropiado. Segundo, el contexto es relevante más que en cualquier otro tipo de política de intervención. La evidencia que proviene del monitoreo administrativo de las aportaciones financieras a las iniciativas individuales incluidas en los pactos dejan claro que éstas funcionan en forma distinta. Dichas iniciativas individuales parecen tener diferentes mecanismos en su funcionamiento. Lo que queremos saber es por qué parece que unos pactos funcionan mejor que otros, qué tanto es esto cierto y por qué sucede. El método utilizado debe brindar respuestas que ayuden a revelar las razones de los resultados diferenciados. En otras palabras, más que en cualquier otra circunstancia, necesitamos un método capaz de proporcionar explicaciones. Todas estas consideraciones mencionadas nos llevan a un enfoque metodológico que considere primordialmente el análisis comparativo entre pactos (que provenga de un análisis comparativo preciso que sea leído en conjunto con mayor evidencia macro-estadística). Al respecto, aun si el estudio recientemente comisionado no es un estudio de evaluación estrictamente hablando, el método seleccionado (trabajo de campo y la interacción directa de los grupos interesados en la política local de los territorios en los que el pacto está activo) parece ser suficientemente adecuado para brindar algunas respuestas reveladoras.

Algo positivo para la evaluación es el incremento de la atención institucional dedicada a construir estadística territorial administrativa y económica a un nivel muy detallado (“sistemas locales” y niveles de consejo). Este tipo de información tiene implicaciones importantes para la evaluación de los pactos territoriales u otras formas similares de promoción del desarrollo local, dado que permite la valuación de los cambios en las áreas objetivo utilizando conceptos estadísticos similares y por lo tanto comparables con aquellos disponibles para territorios más grandes (a nivel regional o provincial). Está claro, cómo se sugirió antes, que la disponibilidad de datos estadísticos territoriales no implica que la evaluación de los efectos de los pactos territoriales pueda ser limitada a la comparación de áreas en las que el instrumento está en funcionamiento con las áreas en las que el instrumento no está en funcionamiento.⁷ Sin embargo, la información estandarizada detallada permite que los diversos estudios informativos puedan

ser utilizados conjuntamente con otros métodos de investigación y ello también puede ofrecer una posible guía para escoger áreas de interés [por ejemplo, áreas donde los estudios de caso dirigidos a un objetivo específico, podrían realizarse sobre instrumentos específicos (como el pacto territorial), o también otro tipo de evaluaciones muy informativas del tipo evaluaciones basadas en áreas].⁸

Qué hemos aprendido hasta el momento: estudios académicos, evaluaciones preliminares, experiencia directa y aprendizaje por monitoreo

La mayoría de los estudios sobre pactos territoriales efectuados en los últimos años ha involucrado la investigación académica independiente, principalmente el trabajo de campo realizado por sociólogos que se han concentrado en el tema de acumulación de capital social.⁹ El objetivo principal de estos estudios ha sido la naturaleza de las asociaciones y el desarrollo de las relaciones basadas en la confianza. Dado que dichos estudios se han desarrollado con métodos diferentes, no se esperaba que ofrecieran conclusiones generales. Sin embargo, todos parecen sugerir que el proceso asociado con el pacto, en los distintos casos examinados, desencadenaron alguna dinámica de capital social, aunque en diferentes formas. Dentro del debate académico de estos temas hay importantes conclusiones que sugieren firmemente a favor de que los pactos territoriales posiblemente afecten el desarrollo local. No obstante, dichos trabajos parecen apuntar a que si el pacto ofreció una oportunidad, ésta tomó a un líder social preexistente para aglutinar la alianza, algo que debe estudiarse más extensivamente.

Los sociólogos encontraron también distintos modelos que se utilizan dentro de la formación de las alianzas, algunos de los cuales son de hecho bastante perversos (éstos observan, en algunos casos, la formación oportunista de alianzas). Debido a que no contamos con este tipo de estudios disponibles para un número suficientemente grande de pactos, no sabemos si las buenas alianzas superan en número a las alianzas oportunistas. Sin embargo, está claro que los resultados muestran que el procedimiento utilizado para seleccionar los pactos tuvo ciertas limitaciones en no desincentivar a las alianzas oportunistas. Hubiera sido útil investigar si la magnitud de los incentivos financieros y el mecanismo de selección de los proyectos jugaron un papel relevante.

A comienzos del año 2000 fue realizada una encuesta en todos los pactos en operación en ese momento (46). Dicha encuesta se efectuó sobre la base de un cuestionario estructurado para empresarios e informantes clave.¹⁰ La encuesta fue parte de un estudio autónomo dirigido con el objeto de revelar la motivación de los actores que comparten el pacto de alianza, accediendo a información directamente de los empresarios involucrados sobre las oportunidades y las necesidades de los territorios. El resultado principal del estudio consistió en

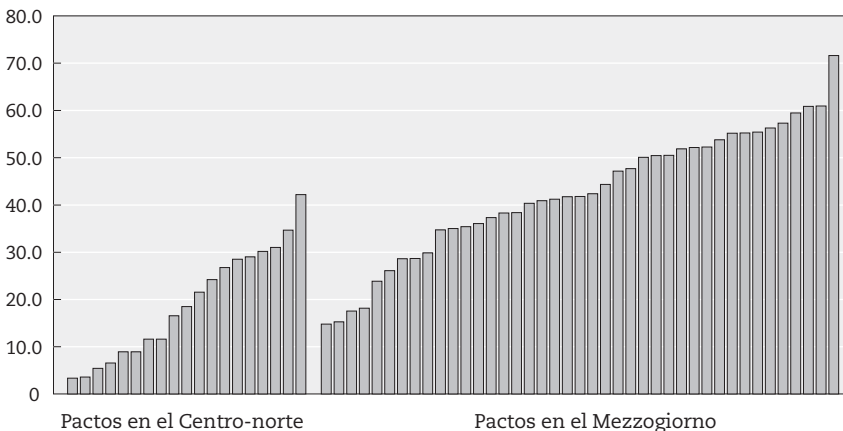
que la política fue particularmente bien recibida en la mayoría de los casos (los empresarios estuvieron en promedio bastante contentos), a pesar de algunas quejas acerca de lo extenso de los procedimientos.

Es más importante, sin embargo, que el estudio produjo una base de datos con información codificada sobre la percepción de los agentes y sus puntos de vista sobre una serie de temas.¹¹ En el mismo año, en el Departamento para el Desarrollo (la autoridad administrativa en ese momento encargada de implementar los pactos) se ejecutó otro proyecto con el fin de armar estadísticas económicas definidas a nivel de pacto territorial, en otras palabras, estadísticas referidas al territorio incluido en el pacto. Dicho ejercicio mostró que los pactos fueron iniciados en territorios con diferentes condiciones económicas, aunque el primer grupo de pactos aparece en áreas con relativamente menores desventajas.

La conclusión más digna de reflexión proviene, casi por casualidad, de la evidencia del monitoreo que se lleva desde finales de 2000.¹² Hasta este punto, el debate general sobre pactos territoriales había considerado a los pactos como homogéneos (los pactos fueron pensados y representados como iguales en todos los lugares, padeciendo de debilidades similares debido principalmente a ineficiencias burocráticas). Las pruebas derivadas del monitoreo mostraron que los pactos estaban actuando de manera diferente respecto al progreso del proyecto del pacto (véase figura 14.2).

Figura 14.2 **Evidencia del monitoreo en la implementación de los proyectos de los primeros 61 pactos aprobados.**

(Centro-norte y Mezzogiorno; porcentaje de la contribución financiera utilizada para cada pacto, datos estandarizados referidos a abril de 2001)



Algunos pactos, después de controlarlos dada su fecha de inicio, revelaron una tasa significativamente más alta de las iniciativas en funcionamiento y una tasa significativamente más alta de gasto. Este tema pareció ser meritorio para mayor investigación.

Tratamos entonces de investigar hasta qué punto las condiciones iniciales de los territorios, medidas por un conjunto simple de variables, habían afectado la formación de alianzas locales, sus objetivos y su funcionamiento.¹³ Estábamos interesados particularmente en descubrir si esas alianzas que parecen desempeñarse mejor (ya que daba la impresión que progresaban más rápidamente en la puesta en marcha del proyecto del pacto) podrían pronosticarse por condiciones iniciales económicas más favorables en los territorios o, en su lugar, habían adquirido y desarrollado las habilidades durante la construcción e implementación del pacto. La idea principal fue observar las diferencias en el desempeño en curso con el fin de revelar las características generales del instrumento durante su funcionamiento, haciendo preguntas de evaluación explícitas (¿Dónde funciona mejor el pacto, por qué? ¿Será debido cuestiones que la política no puede afectar, por ejemplo, condiciones iniciales más favorables? ¿Puede que sea así debido a algo que se pueda agregar para lograr un mejor diseño de la política?)

El estudio utilizó técnicas estadísticas aprovechando un valioso conjunto de datos elaborado sobre la base del monitoreo de la información administrativa acerca de la relativa velocidad en la realización de los pactos (la cual se supone como una representación del desempeño intermedio), características de las áreas de influencia de los pactos y datos provenientes de la encuesta del año 2000.

Dicho ejercicio de investigación se realizó sobre los primeros 61 pactos, poniendo especial atención en aquellos que fueron implementados a través del procedimiento nacional, para lo cual la variable de desempeño fue medida con mayor precisión. Los resultados del análisis pueden ser resumidos de la siguiente forma. Las condiciones económicas iniciales relativas no parece ser que afectaron la implementación del pacto (véase tabla 14.1).¹⁴

Este resultado fue bastante sólido y pudo mantenerse dentro de los varios subgrupos de los pactos examinados. Si las condiciones económicas iniciales no cuentan mucho para explicar los diferentes desempeños, podríamos concluir que es posible fomentar el desarrollo local a través de incentivos para formar alianzas aun en áreas de condiciones económicas bajas.

En cambio, el desempeño de los pactos [respaldado por un ejercicio realizado en un número más pequeño de pactos para los cuales se contó con información disponible adicional (véase tabla 14.2)] pareció estar particularmente relacionado con la calidad del proceso de establecimiento de los pactos a nivel local.¹⁵ Los pactos en los que los actores relevantes estuvieron involucrados en la discusión del

Tabla 14.1 **Implementación de los pactos y condiciones iniciales de los territorios**

PERF	A		A1	
VARIABLES	Coefficiente	(t)	Coefficiente	(t)
Constante	3.301	(0.752)	2.012	(0.494)
DC1	0.547	(1.752)	0.312	(1.036)
DC4	0.455	(1.327)	0.289	(0.900)
LVALAG	-1.242	(-2.286)**	-0.652	(-1.221)
LUNR	0.265	(0.745)	-0.241	(-0.662)
LSIZEFC	0.99	(0.517)	0.092	(0.523)
DS			1.186	(3.242)**
	F(5.55)= 6.04 Prob >F=0.0002 Adj R2=0.2957 Num Obs 64		F(6.54)=7.56 Prob>F=0.000 Adj R2= 0.3995 Num Obs 61	

PERFi=CONSTANTE + b1(DC1) + b2(DC4) + b3(LVALAGi) + b4(LUNRi) + b5(LSIZEFCi). PERF es la variable que representa el desempeño en curso del pacto en abril 2001 (véase explicación en el texto); DC1 es una variable dicotómica (*dummy*) para el primer grupo de pactos territoriales nacionales, los cuales fueron aprobados mucho antes; DC4 es una variable dicotómica (*dummy*) identificando los pactos territoriales aprobados con el procedimiento europeo; LVALAG es el logaritmo del valor agregado per cápita en el área de cada pacto en 1996; LUNR es el logaritmo de la tasa de desempleo en el área de cada pacto en el año 1996 (seleccionada como la fecha en que se dieron las condiciones iniciales); LSIZEFC es el control para la dimensión (logaritmo) del total de la contribución pública disponible para cada pacto. La misma regresión luego se corre sumando una variable dicotómica (*dummy*) DS para el lugar en el sur. Los estimadores Mínimos Cuadrado Ordinarios (OLS) y las variables acotadas son tratados con transformación logística.

Fuente: P. Cassavola y F. Utili (2002) *Promozione di partnership locali per incoraggiare lo sviluppo locale: una valutazione preliminare dei patti territoriali*, Sviluppo Locale, vol. IX, núm. 20.

proyecto desde el inicio procedieron más rápido en la implementación, resultado que fue confirmado por otro trabajo de campo.¹⁶ El resultado es interesante, dado que la forma en que el proceso local está organizado es una variable que podrá verse afectada directamente por la forma en la cual la política sea suministrada en un territorio. Enfatiza la importancia de dar suficiente tiempo y recursos a la fase inicial, que es cuando se diseña una política para promover la formación de una alianza local. Esto también sugiere que una exploración precisa del proceso de formación de la alianza (para prevenir proyectos débiles) puede ser crucial para su éxito futuro. Dicho resultado es particularmente informativo debido a que puede ser generalizado a una familia de instrumentos de política basados en un enfoque para el desarrollo de participación de abajo hacia arriba.

Para obtener mayor información parece necesario aún más trabajo de campo dirigido. Esta es la tarea del estudio implementado por un grupo de investigadores que trabajan para el Departamento de Políticas de Desarrollo. En particular, se les ha solicitado encontrar una definición más apropiada de éxito e investigar con mayor profundidad qué hace que el instrumento funcione mejor.¹⁷

Tabla 14.2 **Implementación de los pactos y las condiciones de calidad iniciales de los territorios**

PERF	B1		C	
	Coefficiente	(t)	Coefficiente	(t)
Constante	-5.3035	(-1.37)	-5.058	(-1.326)
DC1	0.228	(-0.851)	0.206	(-0.737)
LVALAG	0.479	(-0.77)**	-0.054	(-0.896)
LSIZEFC	0.97	(-0.483)	0.054	(-0.272)
DS	1.236	(3.417)***	1.186	(3.294)***
LPARTIC			0.279	(-2.195)
TP_TRUST			-0.99	(-1.014)
TP_INF			-0.076	(-0.674)
	F(4.31)=6.02		F(7.28)=4.35	
	Prob>F=0.0011		Prob>F=0.0023	
	Adj R2=0.364		Adj R=0.401	
	Num Obs 36		Num Obs 36	

PERF=CONSTANTE + b1(DC1) + b2(LVALAG) + b3(LSIZEFC) + b5(DS) + b6(LPARTIC) + B7(TP_TRUST) + b8(TP_INF). En la ecuación las nuevas variables derivan de una encuesta llevada a cabo en un número más pequeño de pactos, haciendo preguntas a los empresarios que participan en los pactos; LPARTIC es una variable que deriva de la respuesta positiva a una pregunta a los empresarios acerca de si en el proceso inicial de montar el proyecto se vieron involucrados todos los actores claves; T_TRUST es una variable derivada de las repuestas a la pregunta a los empresarios acerca del rol del pacto en inducir, mejorar o dificultar las relaciones de confianza entre los actores locales. Representa una porción de empresarios que percibieron positivo el rol del pacto en este asunto; TP_INF se deriva de otro grupo de empresarios que señalaron una mejora en la manera como se dio la infraestructura.

Notas

1. Los recursos para la asistencia técnica son suministrados para la preparación de los proyectos y durante la implementación para apoyar el proceso de coordinación.
2. Entre los dos grupos de pactos hay otras diferencias relevantes, entre las principales se encuentra el creciente papel que ha jugado la agencia local responsable de coordinar la implementación en los Pactos Europeos (una característica que puede ser ampliada a los pactos nacionales) y a la presencia de una mayor variedad de iniciativas (mientras que en los pactos nacionales la contribución pública está dada principalmente para financiar actividades empresariales y en un grado mucho menor para financiar infraestructura pública).
3. La bibliografía referente a distritos industriales destaca la importancia del contexto y de las relaciones locales como la base de diferentes trayectorias de desarrollo y este argumento aparentemente habla a favor de los instrumentos de política tales como los pactos territoriales. Sin embargo, la bibliografía sobre capital social inspirada por otro trabajo (siguiendo las famosas contribuciones de Putnam sobre los orígenes del desarrollo regional italiano) ha llegado frecuentemente a la conclusión de que se requiere de una larga trayectoria histórica para construir el capital social. Algunos investigadores por lo tanto han inferido que no hay lugar para el desarrollador de la política de tratar de acelerar dicho proceso de formación de capital social.

4. Estoy utilizando el término rendición de cuentas de manera no totalmente correcta y algo restrictiva con el fin de apoyar la discusión, y hay que tener en cuenta que la mayoría de las preguntas se refieren al aspecto de los pactos territoriales, por lo que las preguntas son del tipo: ¿se está gastando todo el dinero correctamente? En general, ¿el instrumento está generando el incremento prometido de desarrollo?
5. En el debate político reciente algunos líderes de opinión, comprensiblemente preocupados por el gran número de pactos existentes, han pedido que se decida rápidamente sobre cuáles (de entre los muchos lugares) son los buenos pactos – para continuar apoyándolos- y cuáles son los malos pactos – con posibilidad de cancelarse.
6. En otras palabras un conjunto importante de preguntas corresponde a la relación entre el diseño de política y sus resultados. De hecho las características generales de la política (financiamiento a una alianza con un proyecto) no son suficientes para describir completamente la naturaleza de la intervención. Una serie de otras características del diseño deben ser consideradas: ¿qué tipo de proyectos y con cuáles características?; ¿qué actores son elegibles para presentar el proyecto?; ¿qué tan grande es la aportación pública ofrecida? ¿cómo y cuándo el proyecto es presentado?; ¿cuáles procedimientos deben ser seguidos en la selección de proyectos para financiamiento? Todos estos elementos podrían tener, al menos en parte, un papel en la determinación del éxito futuro de la asociación seleccionada.
7. Existen muchas razones teóricas para desanimar la aplicación mecánica del llamado paradigma de evaluación estándar (el cual compara el grupo tratado con el grupo no tratado) para instrumentos de pactos desarrollo local y sus similares.
8. Evaluaciones basadas en el área podrían ser un método prometedor para la evaluación de las políticas de desarrollo local dado que se requiere que los investigadores comiencen su trabajo en una limitada porción del territorio y desde ese ángulo reconstruir las posibles cadenas causales relacionando los hechos observados y los comportamientos a las intervenciones de la política en esa área.
9. Véase, por ejemplo, el número 2001/3 de la revista italiana *Stato e Mercato*, dedicada en su mayoría a los pactos territoriales.
10. *Sviluppo Italia – Iter (2000) “Caratteristiche e potenzialità dei Patti territoriali”*, Roma.
11. La mayor parte de los estudios sociológicos citados no han producido información codificada o estandarizada (para ser utilizada en otros estudios).
12. Ministerio del Tesoro, Bilancio e Programazione Economica, “Terzo Rapporto sullo sviluppo Territoriale”, noviembre de 2000.
13. Los resultados reportados aquí brevemente provienen de Casavola, P. –Utili, F. *Promozione di partnership locali per incoraggiare lo sviluppo locale: una valutazione preliminare dei patti territoriali*, *Sviluppo Locale*, vol. IX, núm. 20, 2002.
14. Nosotros observamos la relación entre una variable dependiente que describe el desempeño del pacto en un determinado punto en el tiempo (el monto de la aportación financiera pública utilizada en abril de 2001 sobre el monto total disponible en la misma fecha, sobre la hipótesis de que todas las iniciativas fueron progresando a la mayor velocidad) y el indicador proxy a las condiciones económicas de las áreas del pacto (valor agregado per cápita y tasa de desempleo registrada un periodo anterior a la activación del pacto). Unas cuantas variables de control más fueron añadidas.
15. La variable fue recobrada por la encuesta 2000 y refleja qué tan cerca se encontraba de haberse completado; como la percepción de los empresarios en los pactos; del conjunto de actores locales involucrados en la discusión del proyecto del pacto.

16. Dipartimento per le politiche di sviluppo e coesione, *Relazione sui patti Territoriali* (junio 2001).
17. El estudio se realizó durante la segunda mitad de 2002 e involucró el trabajo de campo de 19 casos. El informe final salió en enero de 2003. Los resultados del estudio no son fáciles de resumir sin la pérdida de detalles interesantes. Sin embargo, los investigadores sugirieron una interesante lista de temas que pueden estar asociados con el éxito (abarcando desde la capacidad de las alianzas para proporcionar proyectos innovadores hasta su habilidad de encontrar nuevas fuentes de financiamiento para implementar proyectos no incluidos inicialmente). Una síntesis de la investigación es presentada en líneas en www.dps.tesoro.it/documentazione/docs/patti/RISERCA_PATTI_TERRITORIALI.pdf.

Capítulo 15

Evaluación de los pactos territoriales para el empleo: aspectos prácticos y metodológicos. La experiencia de Austria

*Peter Huber**
Instituto Austriaco de Investigación Económica (WIFO)
Viena, Austria

* El autor agradece a Hedwig Lutz por sus útiles comentarios; los errores son responsabilidad del autor.

Introducción

Los pactos territoriales de empleo son intervenciones en las políticas de los gobiernos locales destinadas a la generación de instituciones y capital social para el mejoramiento de la economía política en ciertas regiones. La premisa que subyace a esta forma de intervención es que las actividades políticas, que se desarrollan en forma autónoma, colectiva y localmente abarcando asociaciones, son propicias para el incremento de la efectividad y la eficiencia en el suministro de una política. En relación a esta premisa, es claro que se debe adoptar una perspectiva empírica en la evaluación de pactos territoriales de empleo. En el mejor de los casos, una evaluación tendría que establecer si los individuos de una región dada se han visto beneficiados con los pactos territoriales de empleo, de qué manera y hasta qué punto. Si los individuos se han beneficiado, entonces se abre la posibilidad de intentar comparar los beneficios con los costos y así derivar un beneficio neto del pacto.

Tal tipo de evaluación, no obstante, conlleva ciertos problemas metodológicos. En particular, en lo que a evaluaciones del mercado de trabajo se refiere, cualquier éxito proveniente de alguna medición específica que generalmente se atribuye a la medición misma y no a la institución que la diseñó. Evaluar la eficiencia de las instituciones para implementar programas implica contar con un contrafactual diferente al utilizado en gran parte de la bibliografía acerca de la evaluación de alguna política activa en el mercado de trabajo. Por ejemplo, en vez de preguntarse “¿qué habría pasado con una persona en particular si él o ella no hubiera sido incluida?”, el contrafactual en este caso ayuda a responder la siguiente pregunta: “¿qué habría ocurrido si la institución que diseñó tal medida no hubiera existido?” Lo que puede hacer muy difícil identificar con precisión al contrafactual.

En este documento se presenta un esbozo del enfoque de evaluación seleccionado para evaluar los pactos territoriales de empleo derivados del Fondo Social Europeo para el Objetivo III (ESF Objective Three Evaluation) en Austria. En particular, en este trabajo se argumenta que, al utilizar conceptos provenientes de procesos de evaluación, los enfoques pueden ser capaces tanto de identificar los logros de los pactos como los impedimentos para su éxito. En

el siguiente apartado de este trabajo se describen algunas características de los pactos territoriales de empleo (TEPs, por sus siglas en inglés) austriacos que son relevantes para una evaluación. En secciones subsecuentes se discuten de manera breve aspectos relacionados con los datos y se destaca el método de evaluación seleccionado. Al final, en otra sección, se presentan las conclusiones.

Las particularidades de los pactos territoriales de empleo en Austria

Aunque existe una larga tradición a través de la cual se coordina la política de mercado de trabajo regional en Austria, los pactos territoriales de empleo son instrumentos de política relativamente nuevos. La iniciativa original proviene de la Comisión Europea que, en 1997, solicitó que algunos proyectos se sometieran a una iniciativa cuyo objetivo era mejorar la situación del empleo. La intención de esta medida era combatir el desempleo a través del diseño de programas específicos que pudieran aglutinarse bajo el título de “programas territoriales de empleo (TEPs, por sus siglas en inglés)”. La Comisión seleccionó cuatro pactos austriacos (Salzburgo, Tyrol, Vorarberg y Viena) para esta iniciativa.

La idea de los pactos territoriales de empleo fue bien recibida y, para 1998, el Ministerio de Empleo, Salud y Asuntos Sociales alentó a las oficinas regionales de servicios públicos de empleo (AMS) y a los gobiernos de las provincias (Bundesländer) para disponer TEPs posteriores. En el marco del plan de acción nacional para el empleo se introdujeron subsidios especiales para respaldar dichos acuerdos territoriales.

Basados en esta iniciativa nacional, los TEPs se desarrollaron rápidamente. Para el año 2002 se estableció un TEP de corte provincial en cada una de las nueve provincias de Austria (es decir, en la actualidad hay nueve TEPs provinciales en ese país). Además, tomando como base las recomendaciones provenientes de un estudio temprano realizado por la OCDE (Campbell, 2001), en un número de pactos, y a través de varios arreglos institucionales, las iniciativas se delegaron a un nivel regional menor.

Los pactos provinciales, en los cuales se enfoca este documento, varían ampliamente tanto en sus metas como en la manera en que definen su propia función. Muchos pactos (por ejemplo, los de Alta Austria, Viena y Baja Austria) ponen particular énfasis en su papel como foro para la coordinación de políticas (en particular, políticas de presupuesto para el mercado laboral activo de PES provinciales y las políticas económicas de los gobiernos de las provincias) tanto en términos de coordinación del presupuesto como de diseño de la política, poniendo menor énfasis en su papel en el diseño de mediciones innovadoras. No obstante, algunos pactos (por ejemplo, en Styria) se definen a sí mismos como una fuente de innovación responsable de diseñar nuevas medidas para el

mejoramiento de la coordinación entre las políticas económicas en su provincia. Los pactos que definen su función principal en términos de instrumento coordinador, como si fuese una regla, adquieren responsabilidades en torno a coordinar la utilización del presupuesto tanto del servicio de empleo público provincial como de los gobiernos provinciales, aun cuando estos recursos de hecho no sean administrados por los pactos.¹ El aspecto central de estos pactos es, por lo tanto, cuál ha sido el “valor agregado” que generan al coordinar este tipo de presupuestos. Asimismo, el grado en que dicha coordinación rebasa al mero presupuesto, así como una discusión general sobre políticas relevantes, varían de pacto en pacto.

Otra particularidad más de los TEPs austriacos es que operan a nivel de provincia, es decir, sobre unidades regionales relativamente extensas (en promedio, alrededor de un millón de habitantes) que a menudo se caracterizan por diferencias internas en las condiciones del mercado laboral. Como ya fue señalado en los primeros análisis de la OCDE (véase Campbell, 2001), esto provee a los pactos de los recursos necesarios y se asegura que los actores que tienen un poder sustancial de decisión se involucren en ellos. Pero los pactos pueden también ser demasiado amplios como para asegurar el involucramiento y la motivación de todos los tomadores de decisiones potenciales.²

Biffl *et al.* (2000) sostienen que, en la mayoría de los casos, los TEPs se establecieron como cuerpos para la coordinación de actividades de las AMS de las provincias y de los gobiernos provinciales. En la mayoría de los TEPs, ambas instituciones fueron socios dominantes. Asimismo, de acuerdo con la misma fuente, había un marcado énfasis en el involucramiento del sector de instituciones públicas y de los socios sociales, mucho más que en el de las instituciones del sector privado incluidas en la asociación.³

A pesar de la heterogeneidad substancial, existen numerosos documentos oficiales publicados tanto por el ministerio como por la oficina de coordinación de pactos austriacos, misma que es responsable de coordinar los pactos territoriales de empleo a nivel federal. Aunque, en términos generales, estos documentos tienden a ser relativamente imprecisos en lo que se refiere a problemas concretos que deben ser abordados, los documentos resaltan el papel que tienen los pactos en coordinar las políticas económica y del mercado de trabajo.

Además, la oficina de coordinación presentó una lista de características comunes para los pactos. En particular, esta lista sugiere que todos los pactos deben estar basados en el análisis de los problemas existentes del mercado laboral, de las estrategias y de las metas compartidas por todos los actores.⁴

Aunque en Austria los pactos son relativamente nuevos, en varias ocasiones han manifestado interés en evaluaciones externas y en la consultoría para futuros desempeños. En la mayoría de los primeros estudios se discute la estructura y

el desarrollo de los pactos. Por ejemplo, respecto al alcance regional óptimo de los TEPs, Campbell (2000) sugirió que debe establecerse una diferenciación analítica entre los niveles estratégico y operativo de los pactos. En relación a los aspectos estratégicos, es imperioso incorporar a los tomadores de decisiones que tengan un poder de decisión relevante y el dominio sobre los recursos adecuados. Esto sugiere que la organización debe darse a un nivel regional más grande. En contraste, el nivel operativo de los pactos debe asegurar el involucramiento, la participación y la motivación de todos los actores locales importantes. Esto se puede llevar a cabo mejor en contextos regionales pequeños, similares a las regiones a nivel NUTS3. En la mayoría de los pactos de las provincias más grandes de Austria se le dio seguimiento a esta sugerencia.

En fecha más reciente, Leitner *et al.* (2002) dirigieron una evaluación detallada y cuidadosa del impacto del pacto territorial de empleo en Viena, y una evaluación todavía en curso del pacto en Styria. Los hallazgos más relevantes de Leitner *et al.* consisten en que las mediciones del TEP son más intensas y mejor dirigidas que otro tipo de mediciones, y que esas mediciones han creado ganancias netas de empleo. Sin embargo, son críticos ante el hecho de que, comparado con los TEPs de otros países, el TEP vienés todavía no está suficientemente focalizado.

Además, la oficina nacional de coordinación (ZSI) ha creado una serie de medidas dirigidas a la autoevaluación. Por ejemplo, la realización de entrevistas a los presidentes de las organizaciones de socios sociales (véanse Scoppeta, 1999 y ZSI, 1999a). Aunque los informes no son acrílicos, por ejemplo, en cuanto a la falta de incentivos para impulsar al empresariado (en el caso de los empleadores) o en relación a la necesidad de crear más iniciativas operativas a nivel local (sindicatos), y aunque en dichos informes también se llame la atención a la necesidad de evaluar los pactos tanto desde una perspectiva individual como de una comparada, todos los socios sociales le dieron la bienvenida a la idea de los TEPs y resaltaron su importancia en la coordinación de políticas locales. En particular, los socios sociales elogiaron los TEPs por su influencia en el incremento de la importancia concedida a la necesidad de coordinar políticas, por su éxito para motivar a un amplio rango de actores regionales y por su habilidad para contribuir al incremento de la flexibilidad en las políticas para el mercado de trabajo.

Situación de los datos relacionados con los pactos

Una particularidad de los pactos territoriales de empleo es que generan sólo unos cuantos datos administrativos. En algunos pactos, aquellos que se relacionan más con la coordinación de las mediciones políticas que con la dirección o el diseño de dichas mediciones, los datos concernientes al presupuesto otorgado para las mediciones del pacto, o sobre las personas que estuvieron involucradas

en las mediciones, entre otros aspectos, son difíciles de obtener y muy poco significativos. Además, este tipo de datos no son recabados a través del pacto sino por los socios. No obstante, los pactos generan cantidades sustanciales de información en forma de textos que, o están obligados a proporcionar, o bien, proporcionan por iniciativa propia. Para los pactos es un requisito proporcionar los siguientes documentos:

- El contrato del pacto. Este contrato se renueva regularmente pero puede extenderse por varios años. Contiene información de los socios del pacto, un análisis de la situación prevaleciente en los mercados de trabajo de la región y los detalles de la estrategia propuesta por los pactos.
- El programa de trabajo del pacto. Contiene información adicional sobre las metas, estrategias y análisis de pactos en los que el contrato no se renueva anualmente.
- La solicitud de subsidios de la ESF. Ofrece detalles sobre los fondos solicitados por alguna ESF y especifica a qué propósitos se destinarán esos fondos, los socios del pacto (dividido en socios que financian, apoyan y ofrecen consultoría) y alguna información acerca de los problemas de la región y las estrategias propuestas en el pacto.

Dado que algunos de estos documentos se producen anualmente, o al menos de manera regular, se trata de textos que permiten observaciones sustanciales acerca de la asociación, respecto de su estado actual así como en términos de su desarrollo en el futuro, las metas fijadas por el pacto y la calidad de los análisis del pacto. Asimismo, varios pactos han publicado información adicional en forma de monitoreos detallados y reportes de implementación, estrategias y estudios que consideran el desarrollo futuro de un pacto (como la integración de estructuras regionales, etcétera), documentos que constituyen fuentes de información adicional para las evaluaciones.

Los pactos suministran estos documentos y son tan heterogéneos como los pactos mismos. Además, existe un reporte estandarizado sobre las actividades del pacto, sus asociados y otros detalles del pacto que la oficina de coordinación de todos los pactos (ZSI) proporciona cada año para todos los participantes.⁵ Adicionalmente, existen otras fuentes de información, como los documentos de política económica publicados por los gobiernos de las provincias y los documentos de los PESs, en todas las provincias.

Esta situación en torno a los datos sugiere que una evaluación de pactos territoriales de empleo –además de estar basada en entrevistas tanto con los socios del pacto como con las personas encargadas de administrarlo– debe tomar en cuenta estas fuentes de información para generar indicadores objetivos del desarrollo de un pacto.

Un enfoque de evaluación

Una evaluación de pactos debe considerar la situación existente de los datos así como también los arreglos institucionales, tales como la extensión y la calidad de la asociación, la forma de definir los objetivos y la manera en la que son implementados en tanto que forman parte del proceso de evaluación. En particular, una evaluación tal tiene que tomar en cuenta que:

1. Los TEPs son sociedades nuevas, que están fuertemente dirigidas a cambiar el comportamiento de actores regionales de tal forma que se pueda proporcionar coordinación en diferentes campos de la política a nivel de análisis, estrategia e implementación. En este contexto, “la calidad de la asociación” es de crucial importancia para el éxito de la medición a largo plazo.
2. Los TEPs están en un estado continuo de desarrollo y aprendizaje. Por lo tanto, este proceso de aprendizaje debe tomarse en cuenta en la evaluación, con la finalidad de entender los errores cometidos y cuándo se corrigen, lo que puede traducirse en precondiciones importantes para el aprendizaje.
3. Los TEPs son una de las muchas instituciones que operan en el mercado laboral y la política de empleo. Por lo tanto, deben considerarse las relaciones con otras instituciones para así obtener una fotografía completa del valor agregado de los pactos. Lo anterior es particularmente importante porque los efectos multiplicadores así como los de sustitución y desplazamiento podrían aparecer entre instituciones.

En principio, en este contexto una evaluación podría elegir concentrarse en varios aspectos de los TEPs. Por ejemplo, se puede tomar la decisión de observar al pacto en tanto institución. En este caso, debería ponerse particular énfasis en las funciones de los actores en la asociación, la naturaleza y el grado de la cooperación, las metas y los procesos de aprendizaje. El objetivo último de tal evaluación sería determinar si los pactos han contribuido en el establecimiento de capital social en la región. La fuerza de este enfoque radica en que está completamente engranado para identificar si es que se consiguió uno de los principales objetivos de los pactos (es decir, la construcción de capital social en la región).

De manera alternativa, la atención puede centrarse en el pacto como programa. Bajo este enfoque el énfasis estaría puesto en tratar de responder preguntas sobre si la estrategia del pacto parece adecuada y cuáles son los resultados probables.

Otra opción es concentrarse directamente en los procesos (véase Schmidt, 1996). En este enfoque, el objetivo de la evaluación es analizar la formulación de la política, su implementación y su asimilación así como los efectos de la política

con la intención de identificar la conexión entre estos elementos. La ventaja de este tipo de enfoque es que se puede abarcar el ciclo de la política completo, al que están sujetos los pactos territoriales.

Formulación de la estrategia

A nivel de formulación de la estrategia, el punto clave es determinar cómo se ha ido desarrollando la estrategia y quién tiene influencia sobre la misma. Por lo tanto, el análisis de la formulación de estrategias debe tomar en cuenta tres aspectos separados:

- Primero, en el contexto de los pactos territoriales de empleo, presuponer que el capital social y la cultura de la cooperación entre actores conduce al mejoramiento de los resultados de la política sugiere que la “calidad de la asociación” debería ser una parte central del análisis. Esta “calidad” de la asociación puede operacionalizarse a través de múltiples indicadores cualitativos y cuantitativos, tales como el nivel de la asociación (número de socios, tipo de socios), que puede tomarse de los documentos del pacto, la presencia de una “cultura de la cooperación” entre los socios, la apertura del pacto a los extranjeros, que puede identificarse a través de entrevistas con personas locales y extranjeras, y el dominio de ciertos socios, mismo que puede establecerse a través de una combinación de indicadores como la ubicación de la oficina del pacto, el reflejo de los objetivos de los socios en los documentos de los pactos, etcétera.
- Segundo, para un análisis de la formulación de estrategias, un elemento importante es determinar hasta qué punto los pactos de hecho han integrado los diferentes campos de la política a un nivel estratégico. Tomando como base los documentos de los pactos (definidos en la última sección) se pueden construir numerosos indicadores respecto de la integración a nivel estratégico. En particular, uno podría determinar en qué grado las estrategias del pacto abarcan aspectos relacionados con más de un campo político y hasta qué punto estos campos políticos están integrados en una sola estrategia coherente.⁶ Asimismo, mediante la observación de las estrategias publicadas por los diseñadores de políticas en torno a varios ámbitos de una provincia, uno puede determinar hasta qué punto la integración ha trascendido el estrecho alcance de los documentos del pacto y ha encontrado aceptación en otras partes. Otro aspecto importante de este análisis es observar el contenido de la estrategia desarrollada por un pacto y determinar su relevancia para los problemas regionales así como la posibilidad de que sea exitosa.
- Tercero, el análisis debe tomar en cuenta los cambios, tanto de la asociación como del contenido de la estrategia, que suceden en el transcurso del tiempo. En este contexto, indicadores tales como el número de cambios de socios puede

proporcionar información valiosa en cuanto a la estabilidad y el desarrollo de la asociación, así como el seguimiento del desarrollo de las estrategias a través del tiempo puede proporcionar información relativa a la flexibilidad del pacto en términos de sus propias estrategias.

Análisis de la implementación

En el caso de la implementación, la atención se centra en los obstáculos para implementar el programa del pacto o su estrategia. En particular, el énfasis está puesto en verificar si el pacto está dotado con los recursos organizacionales y financieros adecuados para instrumentar la estrategia. En este contexto, son tres los pasos necesarios para el análisis:

- Primero, deben considerarse los conflictos que podrían estar obstaculizando la habilidad del pacto para implementar su estrategia. Esto es necesario debido a que los pactos no son organizaciones monolíticas y tampoco operan en el vacío institucional. Por lo tanto, varios conflictos podrían limitar la capacidad de los pactos para implementar su estrategia. Estos pueden clasificarse en dos dimensiones: conflictos con otras instituciones o al interior de los pactos; y conflictos con instituciones sea al mismo nivel regional (horizontal) o a otro nivel (vertical). Se debe otorgar particular énfasis al tema de si los conflictos surgen porque los pactos no cuentan con los suficientes atributos como para implementar programas y hasta qué grado esto ocurre. La información al respecto de estos conflictos puede, por un lado, recolectarse a través de entrevistas. Por otro lado, la información al respecto del potencial para el mejoramiento de una asociación, así como acerca de sus problemáticas, puede obtenerse de un análisis estructural de los actores regionales y de las relaciones que sostienen entre ellos.
- Segundo, al dirigir la atención a la implementación y los conflictos entre socios y con otras instituciones cabe preguntarse: ¿qué habría pasado si los TEPs nunca hubieran existido? ¿Y cómo es que la creación de los TEPs ha impactado en otras instituciones? Tres efectos de importancia particular pueden analizarse. Estos son: a) pérdidas de peso muerto, la posibilidad de que algo del comportamiento observado también hubiese ocurrido en la ausencia de los TEPs. En particular, la vaguedad del conjunto de metas de los TEPs parece sugerir un potencial para tales efectos de peso muerto. Esto puede llevar a situaciones en las que los actores no se sientan limitados por los objetivos de los TEPs y por lo tanto utilicen los subsidios para financiar instituciones establecidas con anterioridad;⁷ b) efectos de dislocamiento y sustitución: se refieren a la posibilidad de que la presencia de los TEPs haya limitado la eficiencia de otras instituciones con tareas similares. Esto puede ser particularmente importante en el caso de pactos que tienen como función principal diseñar mediciones innovadoras y, como consecuencia, compiten

de manera directa con otras instituciones que también se dedican al diseño de dichos proyectos; y c) efectos multiplicadores: son experiencias positivas de pactos territoriales de empleo que podrían haber llevado a la coordinación en y por encima de la coordinación observada en pactos territoriales de empleo. Aunque existen numerosas razones que permiten creer que estos efectos pueden jugar un papel en los pactos territoriales de empleo, es muy difícil cuantificarlos. Hasta cierto punto, observar las dinámicas del desarrollo, en particular en lo que a la formulación de una estrategia se refiere, puede acercarnos al entendimiento de la probabilidad de pérdidas de peso muerto. Además, la realización de entrevistas a profundidad a los socios del pacto, a competidores y a personas responsables de la implementación puede reportar resultados adicionales.

- Tercero, además de analizar cuestiones relacionadas con la interacción entre pactos individuales, en este paso el análisis debe considerar si a los pactos se les dotó con recursos organizacionales y financieros relativos a sus tareas para, de este modo, determinar la eficiencia organizacional de los pactos. En este contexto tiene que determinarse los costos de instalación y operación de los TEPs. Asimismo, información acerca de los recursos financieros disponibles en el presupuesto de los TEPs, así como información adicional sobre los costos de los socios para el pacto, en términos del tiempo empleado en reuniones, etcétera, pueden obtenerse de entrevistas realizadas a los socios del pacto.

Política de asimilación

Los pactos territoriales de empleo están diseñados para cambiar el comportamiento de actores regionales de una manera particular que, en términos generales, podría definirse como destinada a lograr que los actores: a) coordinen y comuniquen las actividades entre ellos; b) desarrollen perspectivas compartidas acerca de los problemas del mercado de trabajo; y c) diseñen una política coherente que tome en cuenta las acciones de cada quien. La cuestión central en el análisis de la política de asimilación es establecer si, y en qué medida, el comportamiento de los actores regionales cambió por los pactos territoriales de empleo.

Para abordar este aspecto es importante considerar la estructura de incentivos de los actores regionales en términos de sus motivos extrínsecos e intrínsecos. En particular, asuntos como la responsabilidad (por ejemplo, quién obtiene crédito por el éxito de los pactos y quién es responsable de las fallas) y la transparencia hacia fuera tienen una importancia primaria en este análisis. Aunque este aspecto ya ha sido tratado con relativa importancia en estudios previos sobre evaluación de pactos territoriales (OCDE, 2001), en los cuales se encontró que la vaga definición de “derechos de propiedad” que resulta de los

pactos lleva a una falta de motivación entre los socios, datos sobre este particular sólo pueden obtenerse a través de entrevistas a los socios.

Asimismo, se puede formular un indicador basado en el comportamiento de la política de asimilación mediante la observación del desarrollo de los pactos por sí mismos. Si los socios consideraron la política de manera positiva se esperaría que, en la arena política, los pactos hayan incrementado sus capacidades. Por lo tanto, observar si los pactos han profundizado sus actividades regionales, ampliado su asociación o diversificado sus contenidos puede proporcionar información adicional acerca de la asimilación de los pactos. Finalmente, una política de asimilación exitosa también implicaría que los pactos cumplen con el criterio de calidad común establecido por la oficina de coordinación de pactos austriacos (véase ZSI, 1999).

Conclusiones

En este documento el propósito es destacar el enfoque elegido para la evaluación de pactos territoriales de empleo en el marco del ESF (Fondo Social Europeo para el Objetivo III) en Austria. En particular, argumenté que el manejo de conceptos que emergen del proceso de evaluación puede resultar de utilidad tanto para identificar logros y obstáculos en el éxito de los pactos, como para ofrecer posibilidades de vislumbrar el desarrollo futuro de estas instituciones. Aunque el método aquí propuesto de alguna forma se dirige a la evaluación de pactos, no está libre de problemas. En particular, el enfoque propuesto encontrará problemas si se exige demasiado a la evaluación en términos de estimaciones cuantitativas de impactos en el mercado laboral.

Notas

1. Generalmente los fondos son manejados mediante un contrato entre socios especificando el uso de fondos de un socio particular para mediciones diferentes. Entonces los socios administran los fondos.
2. En las provincias de Austria las oficinas públicas de servicio de empleo tienen la autoridad para diseñar e implementar medidas apropiadas en los territorios de sus respectivas provincias y coordinar políticas con los gobiernos de las provincias. En contraste, las oficinas de distrito son sólo responsables de la implementación. Además, los gobiernos de las provincias son responsables de desarrollar la estrategia económica y la planeación territorial. Ellos dan cuenta de una parte substancial del gasto. Por debajo del nivel de provincia, las comunidades son los únicos cuerpos administrativos autónomos.
3. Los problemas para involucrar socios del sector privado no son exclusivos de los pactos austriacos. Muchos pactos europeos enfrentan problemas similares (véase por ejemplo las experiencias reportadas en: EC, 1998).
4. El análisis y las estrategias deben existir en forma escrita para todos los pactos y se renuevan en intervalos regulares en numerosas provincias. De hecho, estos documentos sirven como una importante fuente de datos para la evaluación.

5. Este se encuentra disponible en www.pakte.at
6. En principio, pueden imaginarse tres situaciones. Primero, la estrategia podría centrarse en un campo exclusivo de la política (como la planeación activa de medidas de política del mercado de trabajo). Segundo, muchos campos de la política pueden ser dirigidos sin demasiada integración. Finalmente, los campos de la política pueden estar integrados en una estrategia coherente.
7. En realidad, esto se confirma en la evaluación reciente del pacto de Viena (Leitner et al., 2002) donde un encuestado respondió a la pregunta de si las metas difusas son un problema diciendo: “En mi percepción, esto nunca fue un problema. La ventaja es que puede implementarse más fácilmente, la desventaja es que el resultado es el mismo al ya obtenido” (Leitner et al., 2002).

Referencias bibliográficas

- AMS WIEN and WAFF (2000), Territorialer Beschäftigungspakt Wien 2000, Viena.
- BIFFL et al (2000), Evaluierung des Nationalen Aktionsplanes für Beschäftigung in Österreich, WIFO, Viena.
- CAMPBELL, Mike (2001), “Partnerships in Austria: Enhancing Regional Co-operation in a Decentralised Policy Framework”, en OCDE (2001).
- EUROPEAN COMMISSION (1998), “Territorial Employment Pacts: Good Practice Seminar, Bremen 25-27 January 1998”, Informe del Seminario.
- EUROPEAN COMMISSION (1999), Leitfaden der Europäischen Kommission für die territorialen Beschäftigungspakte 2000-2006, www.pakte.at/rahmendokumente.html.
- GEYR, Renate and Erika HESS (2000), “Der arbeitsmarktpolitische Verbund” 12/23, Viena.
- HUBER, Peter and Ewald WALTERDKIRCHEN (1999), “Möglichkeiten Regionaler beschäftigungspolitik in Oberösterreich”, Studie des Österreichischen Instituts für Wirtschaftsforschung im Auftrag des Landes Oberösterreich, WIFO, Viena.
- KATAJAMÄKI, Hannu (1998), Beginning of local Partnership in Finland – Evaluation, interpretation and impressions, Research Institute at the University of Vaasa, Publication núm. 76, Vaasa.
- LEITNER, Josef (2000), Machbarkeitsstudie – Schaffung von Arbeitsplätzen innerhalb von Projekten auf kommunaler und (klein) regionaler Ebene, www.pakte.at/rahmendokumente.html.
- LEITNER, Andrea, Angela WROBELSKI, Peter PRENNER, Helmut HOFER, Andreas SCHUH, Helmut MAHRINGER (2002), Evaluation der Arbeitsmarktpolitischen Maßnahmen des TBP Wien 1999, Institute for Advanced Studies, Viena.
- TERRITORIALER BESCHÄFTIGUNGSPAKT SALZBURG (2000), Gesamtbericht – Arbeitsmarktpolitische Initiativen des Landes Salzburg 1997-1999, Salzburg.
- OECD (2001), Local Partnerships for Better Governance, París.
- PAKT FÜR ARBEIT UND WIRTSCHAFT TIROL, PAKT FÜR ARBEIT UND WIRTSCHAFT TIROL – RAHMENVERTRAG, www.tirol.gv.at/.

- PAKT FÜR ARBEIT UND WIRTSCHAFT TIROL, PAKT FÜR ARBEIT UND WIRTSCHAFT TIROL, Innsbruck.
- PAKT FÜR ARBEIT UND WIRTSCHAFT TIROL, ARBEITSMARKTVEREINBARUNG LAND TIROL – ARBEITSMARKTSERVICE TIROL, www.tirol.gv.at/.
- SACHSE, Irene, Klaus WIRTH and Gerda ZEMAN-STEYRER (2000), “Kommunale Beschäftigungspolitik in Österreich Status Quo – Best Practices”, Studie im Auftrag des Bundesministeriums für Arbeit, Gesundheit und Soziales und des ESF, www.pakte.at/rahmendokumente.html.
- SAIKKONEN, Paavo, “Observations and Comments on the Austrian Peer Review from the Finnish viewpoint”, manuscript, Ministry of Labour, Helsinki, Finlandia.
- SCOPPETTA, Anette (1999), “Territoriale Beschäftigungspakte – Vernetzung von Politikbereichen auf regionaler Ebene”. In Eichmann, Hochgerner, Nahrada (ed.): Netzwerke – Kooperation in Arbeit, Wirtschaft und Verwaltung, Wien: Falter Verlag, Soziale Innovation + Neue Soziologie, Band 5, S109-126.
- SCOPPETTA, Anette (2000), “Hochkonjunktur für regionale Partnerschaften – Territoriale Beschäftigungspakte im Aufwind”. In LEADER Magazin Österreich. Nummer 2_2002, Wien, 2002, 29-30.
- ZSI (1999), Leitfaden der Territorialen Beschäftigungspakte in Österreich 2000-2006, www.pakte.at/rahmendokumente.html.
- ZSI (1999), Fragen und Stellungnahmen zu den Territorialen Beschäftigungspakten, Viena.

Capítulo 16

Comentario sobre el taller “Evaluación de los pactos territoriales para el empleo: aspectos prácticos y metodológicos”

por
Hugh Mosley
Centro de Investigación en Ciencias Sociales
Berlín, Alemania

Los Pactos Territoriales para el Empleo (PTE) son redes circunscritas en áreas para la coordinación e implementación de políticas económicas y de empleo en proyectos integrados. En Austria e Italia, esta propuesta primero se introdujo en calidad de programa piloto de la Unión Europea; subsecuentemente continuó, con algunas modificaciones, como elemento de política nacional. Es importante observar que los PTE no son organizaciones en un sentido formal, por ejemplo con presupuestos y empleados propios, sino asociaciones (más o menos) voluntarias en las cuales las organizaciones que las constituyen conservan su identidad individual. Sólo se destina una cantidad relativamente pequeña de financiamiento para cubrir los costos para el desarrollo y la coordinación de los programas PTE, lo cual potencia una cantidad desproporcionadamente grande de actividad conjunta de las organizaciones integrantes.

Las redes PTE mismas son extremadamente variadas en y entre países: se diferencian, por ejemplo, en las áreas cubiertas [provincias enteras (Austria) o redes locales (Italia)]; en la constelación de actores públicos y privados involucrados; en la determinación de su función principal [coordinación y planeación (Austria) u operación de programas locales (Italia)] y de su orientación [más hacia el empleo (Austria) o al desarrollo económico (Italia)].

Los PTE poseen dos principios particulares que es necesario considerar para el diseño de una evaluación: 1) se asume que producen efectos de sinergia con la creación de una coalición circunscrita, local y formalizada, poseedora de metas comunes y proyectos coordinados. Si no se dispone de recursos de programas adicionales, lo anterior sólo puede alcanzarse a través del incremento de la eficiencia y la eficacia mediante la aglutinación de los recursos existentes, por ejemplo, a través de una mejor coordinación de las estrategias locales para el empleo. Dirigir los recursos limitados a un menor número de objetivos prioritarios también puede mejorar los resultados de la política. 2) Otro principio importante es el fomento del desarrollo de capital social bajo la forma de redes locales alrededor de esta política (*local policy networks*), que se supone tienen un impacto favorable a largo plazo sobre el empleo y el desarrollo económico locales.

Los artículos presentados por Paola Casavola y Peter Huber, relativos a la evaluación de los pactos territoriales para el empleo en Italia y Austria,

respectivamente, están basados en evaluaciones en curso de programas nacionales. Ambos enfatizan el objetivo de promover redes locales para la implementación de políticas económicas y de empleo; también buscan desarrollar una estrategia de evaluación apropiada, distinta a la del programa dominante de evaluación. Como Peter Huber argumenta: "Evaluar la eficiencia de las instituciones en la implementación de programas implica un contrafactual distinto al investigado en gran parte de la bibliografía sobre política activa para el mercado laboral". En lugar de preguntar qué habría sucedido si un individuo no hubiera participado, "aquí el contrafactual debería responder a la pregunta: ¿qué habría sucedido si la institución que diseña los indicadores no hubiera existido?".

En su estudio de los pactos territoriales en Italia, Paola Casavola proporciona una evaluación cuantitativa de los determinantes de la formación de pactos con base en los datos de 61 distritos italianos. Los supuestos generales del programa italiano de incentivos para la formación de pactos territoriales son que el empleo y el desarrollo económico locales son fomentados 1) por la promoción de densas redes sociales y de cooperación económica y 2) porque existe un efecto positivo de "aglomeración" al combinar los programas individuales en un proyecto de política integral y más amplio.¹ Sin embargo, estos interesantes supuestos no son el tema de su artículo. En cambio, su evaluación para el Ministerio de Economía y Finanzas italiano se centra en las metas intermedias que son el verdadero objetivo del programa italiano: el progreso de los proyectos integrados y las características de las coaliciones locales. Combinando datos de seguimiento del desempeño de los pactos con indicadores sociales y económicos y datos de encuestas sobre el desarrollo de las coaliciones de los pactos, examina varias cuestiones clave para los 61 pactos territoriales italianos: 1) ¿En qué medida la formación exitosa de un pacto es afectada por las "condiciones iniciales", es decir por las condiciones sociales y económicas en las localidades del pacto? 2) ¿Cómo afecta a los resultados del pacto la calidad del proceso de arranque? En ambos casos, utiliza como la variable dependiente la "velocidad de realización", medida en términos del número de trámites para disponer de los fondos autorizados del pacto. En pocas palabras, concluye que la realización de los pactos no fue afectada adversamente por las condiciones iniciales, es decir, esta forma de gobernanza de la política también es factible en localidades en condiciones de desventaja, y que los "pactos" en los que todos los actores relevantes estuvieron involucrados desde el principio avanzan más rápido, lo cual sugiere que la inversión en la fase de arranque es particularmente importante para el éxito de los pactos territoriales.

El artículo de Peter Huber versa sobre el diseño de evaluación y los resultados iniciales de una evaluación de los pactos territoriales para el empleo en Austria, como parte de una evaluación más amplia de los programas en Austria derivados

del Fondo Social Europeo para el Objetivo III, llevada a cabo por el Instituto para la Investigación Económica de Austria (WIFO, por sus siglas en alemán). También Hubert se centra más en los procesos que en los resultados individuales o de mercado laboral. Su pregunta de investigación primordial es un contrafactual institucional: ¿cómo habrían diferido las políticas de las organizaciones participantes en ausencia de la organización del PTE?

El diseño de evaluación se centra en tres tipos principales de cuestiones: 1) formulación de la estrategia, lo cual incluye aspectos relacionados con la calidad de la coalición del pacto y su éxito para alcanzar una integración estratégica de los socios; 2) análisis de la implementación, con particular atención en los conflictos al interior de las coaliciones y el impacto del PTE sobre los actores institucionales participantes, así como en los posibles "efectos multiplicadores" de largo plazo del "capital social" creado por la experiencia del PTE. Huber demuestra la relevancia de las categorías tradicionales del programa de evaluación (*v. gr.*, peso muerto, desplazamiento) incluso en un estudio de caso más cualitativo del impacto de las instituciones de gobernanza. 3) Finalmente, conforme se adopta la política [*under policy-uptake*], aboga por dar particular atención a los incentivos para que los actores regionales entablen coaliciones locales de la misma, lo cual depende de los costos y beneficios percibidos por los socios potenciales de la coalición.

La discusión en torno a los artículos presentados en la conferencia de Viena estuvo dedicada, en buena medida, a cuestiones informativas sobre los programas italiano y austriaco (*v. gr.* diferencias entre los pactos en Italia y Austria y los tipos de programas para el mercado laboral implementados) y aspectos técnicos del análisis (*v. gr.* cómo se mide el éxito). Temas más generales relacionados con el enfoque distintivo de ambos artículos fueron las "cuestiones de gobernanza", en lugar de los resultados individuales del programa, y la dificultad para definir y medir el capital social. Ambos programas, el austriaco y el italiano, están basados en supuestos relativos a la importancia de las redes incluyentes para el desarrollo del empleo y la economía locales, los cuales no se tratan directamente en ninguno de los artículos.

Ambos artículos ilustran la necesidad de una estrategia de evaluación diferente cuando se evalúa el impacto de las estructuras de gobernanza locales. La metodología estándar para la evaluación en la investigación del mercado laboral se ha centrado ampliamente en los efectos individuales de cada programa. Basados en un modelo explicativo tomado en préstamo de la investigación médica, este tipo de evaluación ha buscado estimar el impacto de los "tratamientos" del programa sobre los participantes individuales. En este paradigma metodológico, para estimar los efectos netos, se comparan el empleo y los ingresos de los participantes en el programa con los de un grupo de control experimental o cuasi-experimental de no-participantes (el contrafactual).²

Esta tradición de análisis de impacto ha incrementado enormemente nuestra sofisticación para evaluar los efectos netos de los programas para el mercado laboral. Sin embargo, este modelo estándar posee dos grandes deficiencias desde el punto de vista de la evaluación de las políticas para el desarrollo del empleo y la economía locales:³

1. Tiende a asumir un modelo simplificado del proceso de las políticas, según el cual los "tratamientos" del programa están altamente estandarizados, subestimando la importancia de la variación local del mismo. El "tratamiento" real que reciben los participantes en el programa puede variar considerablemente para participantes en distintos lugares del país. De hecho, los programas para el mercado laboral ofrecen servicios complejos que rara vez son uniformes en su implementación local.⁴ Más aún, los programas para el mercado laboral en muchos países proporcionan meros marcos de regulaciones que por su diseño permiten mucha variación local en el contenido del programa. Éste es el caso, por ejemplo, para muchos programas del Fondo Social Europeo, del cual los pactos territoriales para el empleo austriacos son un ejemplo, así como para los programas más amplios de capacitación y creación de empleo en Alemania y en otros países con servicios públicos para el empleo relativamente descentralizados.⁵ En suma, los "tratamientos" del programa rara vez son tan estandarizados como sugiere la analogía con la investigación médica (Scheirer 1994). Los efectos locales en la evaluación de programas pueden en principio examinarse si el diseño de evaluación proporciona hallazgos válidos en los ámbitos regionales y locales relevantes.⁶

2. Más importante, el énfasis sobre los efectos individuales del programa descuida una gama de otras cuestiones de la evaluación que son centrales para el análisis de la política (evaluaciones de la gestión y los procesos, estudios de eficiencia, análisis costo-beneficio, etcétera) para los cuales otras metodologías pueden ser más adecuadas (véase de Koning y Mosley 2001; Palumbo y Calista 1990; Scheirer 1994). Por ejemplo, cuestiones estructurales relacionadas con el impacto de los cambios en la organización del servicio público para el empleo, incluyendo el financiamiento, la gobernanza y las estructuras de gestión, no pueden abordarse en un marco experimental o cuasi-experimental. De manera semejante, las cuestiones de evaluación relacionadas con el impacto de las estructuras de gobernanza locales, como los pactos territoriales considerados aquí, requieren un diseño de evaluación diferente.

¿En qué es diferente la evaluación de los pactos territoriales?

1. Aunque en ambos países los pactos territoriales para el empleo se originaron por una iniciativa de la Unión Europea posteriormente incorporada en las políticas nacionales, la dimensión "local" es en primera instancia el énfasis territorial o de área de la política, en contraste con las políticas orientadas a los individuos o a las empresas. La mejora en los programas individuales y en los resultados

agregados se busca a través de la influencia sobre el comportamiento colectivo en la formulación e implementación de políticas en el ámbito local. El objetivo es transformar las estructuras de implementación locales mediante la promoción de amplias coaliciones para la formulación e implementación de las políticas. Así, la pregunta central de la evaluación aquí es si se alcanza este efecto local y en qué medida, y no, en primera instancia, el impacto de los "tratamientos" del programa sobre los participantes individuales.⁷

2. Una segunda cuestión de la evaluación es el impacto de los programas operados por los pactos territoriales sobre los programas para el mercado laboral local. ¿La nueva forma de gobernanza (PTEs) ha tenido un impacto sobre el nivel y la combinación de política local (para el mercado laboral) en el ámbito local? ¿Los actores involucrados alteran sus prioridades y sus portafolios de políticas como resultado de su participación en la coalición del PTE? Un principio fundamental para este tipo de estrategia de implementación es incrementar la eficiencia y la eficacia de la política a través del mejoramiento de la coordinación entre diferentes agencias y entre actores públicos y privados involucrados en la implementación local de la política. Por ejemplo, en muchos países, diferentes organizaciones son responsables de la política para el mercado laboral y de las políticas para el desarrollo económico local. Esto parece ser una función fundamental en el modelo PET austriaco, que en muchos casos representa acuerdos entre las oficinas regionales del servicio público de empleo y las autoridades de la provincia responsables del desarrollo económico.
3. Otra cuestión distintiva de la evaluación es el impacto de los cambios en la gobernanza inducidos por el programa sobre los resultados agregados de la política en las regiones. Para responder a esta interrogante se requiere una metodología diferente al método de grupo de control para evaluar el impacto de los programas sobre los participantes. Una debilidad de este método, incluso en el caso de la evaluación del programa, es que no toma en consideración el desplazamiento, la sustitución y otros efectos indirectos a nivel agregado en el mercado laboral. Un análisis agregado del impacto a nivel regional puede utilizarse para evaluar los efectos en el nivel local, medidos en términos de indicadores del desarrollo del empleo y la economía locales. Podría utilizar un marco de series de tiempo o de corte transversal (o ambos), o una comparación emparejada [*a matched comparison*] de unidades regionales.⁸ Un análisis agregado del impacto también podría ser, por ejemplo, el marco apropiado para probar el supuesto común de estos programas: mejorar el "capital social" bajo la forma de coaliciones locales para la política induce un mejor desempeño económico local y del mercado laboral.
4. Una cuestión final de la evaluación, la última pero no la menos importante, es la interrogante clásica respecto al impacto del programa sobre los

participantes individuales. Dada la heterogeneidad de los pactos territoriales y el supuesto de que la implementación local es relevante, se requiere una evaluación centrada localmente, idealmente basada en datos sobre los resultados en el mercado laboral local de los participantes y un grupo de control. En principio, la evaluación de programas locales puede llevarse a cabo utilizando los mismos métodos experimentales o cuasi-experimentales aplicados a los programas nacionales, aunque esto parece ser poco frecuente. Por ejemplo, Leitner et al. (2002) examinan el impacto de las medidas de la política para el mercado laboral del Pacto Territorial para el Empleo en Viena utilizando un método cuasi-experimental basado en un procedimiento de probabilidad de participación en el emparejamiento [*propensity score matching procedure*]. Los datos sobre los participantes en el programa y un grupo de control de no-participantes fueron obtenidos de los registros de la seguridad social y utilizados para evaluar el impacto neto de los programas en Viena sobre el empleo y los ingresos de los participantes.

¿Por qué, en este caso, fue posible una evaluación de la participación en los programas locales para el empleo utilizando un método de grupo de control?

- 1) El Fondo Social Europeo ordena la evaluación de sus programas, aunque la calidad real de las evaluaciones puede variar significativamente;
- 2) las economías de escala que hacen tan costosa la evaluación de programas más pequeños en algunas localidades, no fueron un obstáculo en Viena;
- 3) la habilidad de para evaluar estuvo disponible [Instituto para Estudios Avanzados (IHS, por sus siglas en alemán)], lo cual debería ocurrir en la mayoría de los países europeos;
- 4) el elemento clave parece haber sido la disponibilidad inmediata de los datos a partir de registros de la seguridad social sobre las trayectorias laborales de los participantes y no-participantes locales, lo cual no ocurre en muchos países europeos;
- 5) también vale la pena hacer notar el hecho de que el reporte de la evaluación final de los programas locales de Viena para 1999 fue completado la primavera de 2002, lo cual es indicativo de la disyuntiva entre la calidad de la evaluación y la necesidad de retroalimentación en tiempo real de los administradores del programa.

Notas

1. El modelo implícito para el primer supuesto es la historia del éxito de los distritos industriales italianos, el cual es atribuido a su dotación del "capital social" especial.
2. Para una visión panorámica véase, por ejemplo, el artículo de Smith en este volumen.
3. Véase también Mosley, H. y E. Sol (2001).
4. Mosley y Degen (1994) discuten la variación real en servicios proporcionados incluso en los programas de capacitación altamente estructurados del Reino Unido.
5. Schierer (1994) distingue entre programas "agregados" y "dirigidos".

6. La varianza en los efectos netos del programa a nivel local también podría servir como medida de la (inobservada) variación en la implementación.
7. De hecho, los artículos de Peter Huber (Austria) y Paola Casavola (Italia) suscitan la cuestión de qué queremos decir con evaluación "local", que se refiere no sólo a la evaluación de programas locales sino también a efectos locales específicos.
8. Véase de Koning (2001) para una discusión general de los enfoques en el análisis de impacto agregado.

Referencias bibliográficas

- DE KONING, Jaap y Hugh MOSLEY (eds.) (2001), *Active Labor Market Policy and Unemployment*, Cheltenham, Reino Unido: Edward Elgar.
- DE KONING, Jaap (2001), "Models for Aggregate Impact Analysis of Active Labor Market Policy" en Koning y Mosley, 2001.
- LEITNER, Andrea, Angela WROBLEWSKI, Peter PRENNERM, Helmut HOFER, Andreas SCHUH, Helmut MAHRINGER (2002), *Evaluation der arbeitsmarktpolitischen Maßnahmen des TBP Wien 1999*, Viena: Institut für Höhere Studien.
- MOSLEY, Hugh G. y Christel DEGEN (1994), "The Re organisation of Labor Market Policy: Further Training for the Unemployed in the United Kingdom". WZB Discussion Paper FS I 94-205, Wissenschafts-zentrum Berlin für Sozialforschung.
- MOSLEY, Hugh y ELS SOL (2001), "Process Evaluation of Active Labor Market Policies and Trends in Implementation Regimes" en Koning and Mosley 2001.
- PALUMBO, D.J. y D.J. CALISTA (1990), "Opening up the Black Box: Implementation and the Policy Process", en Palumbo and Calista (eds.) *Implementation and the Policy Process: Opening up the Black Box*, Greenwood Press, Nueva York, pp. 3-16.
- PAWSON, Ray y Nick TILLEY (1997), *Realistic Evaluation*, Londres: Sage.
- SCHEIRER, M.A. (1994), "Designing and Using Process Evaluation", en J. Wholey, H. Hatry, K. Newcomer (eds.), *Handbook of Practical Program Evaluation*, San Francisco: Jossey-Bass Publishers, pp. 40-68.

Capítulo 17

Una reseña de las metodologías de evaluación de impacto de los programas de desarrollo microempresarial

por
Gary Woller,
*Profesor Asociado del Instituto Romney de Administración Pública,
Brigham Young University Utah, Estados Unidos*

Introducción

Desde hace varios años, los gobiernos de la OCDE han invertido millones de dólares en los programas de desarrollo empresarial tanto de países miembros, como de los menos desarrollados. (LDCs, por sus siglas en inglés). El desarrollo microempresarial se basa en un par de premisas subyacentes: 1) el auto empleo es un componente clave para crear oportunidades económicas para las personas de bajos ingresos con opciones limitadas para obtener empleo e ingresos de otra manera, y 2) las restricciones fundamentales para el autoempleo productivo entre las personas de bajos ingresos son el acceso a capital (préstamos) y a la capacitación. Para los gobiernos de la OCDE, el autoempleo genera diversas opciones de política para combatir la pobreza en sus muchas manifestaciones. Para el pobre, el autoempleo amplía el rango de opciones para la autosuficiencia y la supervivencia. Las políticas de autoempleo parecen ofrecer beneficios particulares en las economías caracterizadas por un desempleo o subempleo crónico, o por altos niveles de participación de los pobres en la economía informal.¹

En teoría, los participantes del programa utilizan los préstamos y la capacitación como palanca para iniciar y expandir las micro y pequeñas empresas, por lo tanto, generando mayores niveles de retornos empresariales; los mayores retornos empresariales se traducen en mayores ingresos para el hogar; y a su vez el mayor ingreso del hogar es invertido en bienestar socioeconómico para la familia. En la práctica, los impactos específicos de los programas para el desarrollo microempresarial son difíciles de identificar y aún más difíciles de medir. Las evaluaciones del impacto requieren la adopción de metodologías de investigación capaces de aislar impactos específicos del interior de una complicada red de factores causales e intermediarios y altos volúmenes de “ruido” ambiental aleatorio, así como asignar unidades específicas de medida a impactos tangibles e intangibles que pueden prestarse o no a una definición o medición precisa. Esto no es una tarea fácil.

Empero, el desarrollo microempresarial compite con otras políticas de desarrollo del empleo por los escasos recursos públicos, y es razonable que los diseñadores de políticas deberían querer saber si el desarrollo microempresarial es

una buena inversión en relación con otras opciones de política. Afortunadamente, sí existen metodologías para evaluar el impacto del programa. Sustentándose en los principios y experiencia de las ciencias sociales y naturales, las metodologías para la evaluación del impacto (EI) están bien desarrolladas y son bien conocidas por los investigadores. Sin embargo, las mismas metodologías son menos conocidas por los diseñadores de políticas, con las implicaciones que esto pueda tener para el análisis de la política.

La validez de los hallazgos de cualquier evaluación de impacto está en proporción directa con la validez de la metodología de la EI utilizada. Existe una substancial variación en las metodologías utilizadas por los investigadores de la EI, pero no todas las metodologías para la EI son igualmente válidas. Como resultado, la calidad de los estudios de la EI recorren el espectro. Las variaciones metodológicas reflejan diversos factores, tales como la habilidad e inclinación del investigador, el propósito de la evaluación y las restricciones ambientales o de recursos. La realidad es que la mayoría de los investigadores de la EI trabajan bajo restricciones que les requieren hacer compromisos entre la precisión metodológica y la factibilidad metodológica. La validez de un estudio de EI particular tiene a menudo como foco principal la validez de los compromisos aplicados.

Dados los retos metodológicos que aparecen inevitablemente durante cualquier evaluación de impacto, la habilidad de los diseñadores de políticas para lograr decisiones informadas sobre el impacto de las políticas de autoempleo (o cualquier política pública, para el caso) se puede afirmar que depende en gran medida de su habilidad para hacer juicios informados sobre la validez de las metodologías de evaluación utilizadas y los compromisos derivados de ello. Para profundizar a este respecto, este estudio examina las metodologías de EI utilizadas en 67 estudios de EI en 90 programas de microfinanzas en 31 LDCs y en 20 estudios de EI en 20 programas para la microempresa en 2 países de la OCDE (19 de ellos en los Estados Unidos, véanse las Tablas 17.1 a 17.3).²

La siguiente sección describe los fundamentos conceptuales de la EI, seguido por la exposición de los dos paradigmas metodológicos dominantes y los cinco enfoques metodológicos que caen dentro de ellos. Las secciones cuarta y quinta discuten, respectivamente, las principales dificultades metodológicas que afligen la EI y otros defectos metodológicos misceláneos comunes en los estudios de EI. La sección sexta ofrece reflexiones adicionales para juzgar el rigor metodológico de los estudios de EI. La sección final ofrece algunas recomendaciones de política.

Antes de avanzar más, una aclaración viene a la orden. De aquí en adelante, el término *microfinanzas* es utilizado para designar el desarrollo microempresarial en los países menos desarrollados (LDCs), y el término *microempresa* para designar lo mismo en los países de la OCDE. La terminología diferente refleja distinciones importantes entre ambos. En los LDCs, la industria de las microfinanzas está

evolucionando en una industria de servicios financieros completamente desarrollada para los pobres. De manera creciente, las instituciones de microfinanzas (MFIs) están ofreciendo una variedad de servicios financieros —tales como cuentas de ahorro, préstamos para el consumo o seguros— además de los préstamos empresariales. Históricamente, los pobres en los LDCs no han tenido acceso a los servicios financieros formales de cualquier clase. No obstante, las MFIs están descubriendo una gran demanda latente por un conjunto diversificado de servicios financieros formales entre los pobres, y están evolucionando para cubrir la demanda. Entre las metas principales de muchos defensores de las microfinanzas está la “penetración financiera”* o la creación de un sistema financiero de intermediación sustentable para los pobres.

Tabla 17.1 Resumen de la evaluación de impacto (EI) reseñadas y analizadas

Evaluación de impacto reseñada ^a	Año	País	Metodología ^b	Grupo de control ^c	Marco temporal ^d	Comentario ^e
LDCs						
Ashe y Parrot	2001	Nepal	S, Q	N	CS	
Barnes	2001	Zimbabue	S, Q	Y	L	
Barnes, Morris y Gaile	1999	Uganda	S, Q	Y	L	
Bolnick y Nelson	1990	Indonesia	S	Y	CS	
Buckley	1996a	Kenya	S	Y	CS	
Buckley	1996b	Malawi	S, Q	Y	CS	
Buvinic, Berger y Jaramillo	1989	Ecuador	S	Y	L	
Chen y Snodgrass	2001	India	S, Q	Y	L	
Churchill	1995	Sudáfrica	S, Q	Y	CS	
Coleman	1999	Tailandia	S, Q	Y	L	SB, DS, CL
Coleman	2001b	Tailandia	S, Q	Y	L	SB, DS, CL
Copestake, Bhalotra y Johnson	2001	Zambia	S, Q	Y	CS	
Creevey, Ndour y Thiam	1995	Guinea	S, Q	Y	CS	
Deardon y Khan	1994	Bangladesh	S	Y	L	
Diagne	1998	Malawi	S	Y	L	F
Dunn y Arbuckle	2001	Perú	S, Q	Y	L	
Goetz y Gupta	1996	Bangladesh	Q	N	CS	
Gupta y Davalos	1993	Jamaica	S	N	CS	
Hashemi, Schuler y Riley	1996	Bangladesh	S, Q	Y	L	SB, DS, CL
Hulme, Montgomery y Bhattacharya	1996	Sri Lanka	S	Y	CS	
Karlan y Alexander	2002	Perú	S	Y	CS	SB, DS
Kevane y Wydick	2001	Guatemala	S	Y	CS	
Khandker	1996	Bangladesh	S	Y	CS	SB, DS, CL
Khandker	2001	Bangladesh	S	Y	L	SB, DS, CL
Khandker, Samad y Khan	1998	Bangladesh	S	Y	CS	SB, DS, CL
Kilby y D'Zmura	1985	Brasil	S	N	CS	BC, CL
		Alto Volta	S	N	CS	BC, CL
		Honduras	S	N	CS	BC, CL
		Rep. Domin.	S	Y	CS	BC, CL
		Perú	S	N	CS	BC, CL
		Perú	S	N	CS	BC, CL

* En México extensivamente llamada “bancarización” pero cubriendo servicios más allá de la banca.

Tabla 17.1 Resumen de EI reseñadas y analizadas (cont.)

Evaluación de impacto reseñada ^a	Año	País	Metodología ^b	Grupo de control ^c	Marco temporal ^d	Comentario ^e
Lapar, Graham, Meyer y Kraybill ^f	1995	Filipinas	S	Y	CS	SB
Lapar, Graham y Meyer ^f	1995	Filipinas	S	Y	CS	SB
McKernan	1996	Bangladesh	S	Y	CS	SB, DS
MkNelly y Dunford	1999a	Bolivia	S, Q	Y	L	SB
MkNelly y Dunford	1999b	Ghana	S, Q	Y	L	SB
MkNelly, Watetip y Lassen	1996	Tailandia	S, Q	Y	CS	
Montgomery, Bhattacharya y Hulme	1996	Bangladesh	S, Q	Y	CS	
Morduch	1998	Bangladesh	S	Y	CS	SB, DS, CL
Moseley	1996a	India	S	Y	CS	
Moseley	1996b	Indonesia	S	Y	CS	
Moseley	1996c	Bolivia	S	Y	CS	
Moseley	2001	Bolivia	S, Q	Y	CS	
		Bangladesh	S	Y	CS	
		Bolivia	S	Y	CS	
		India	S	Y	CS	
Moseley y Hulme	1998	Indonesia	S	Y	CS	
		Kenya	S	Y	CS	
		Malawi	S	Y	CS	
		Sri Lanka	S	Y	CS	
Mustafa, Ara, Banu, Hossain, Kubir, Moshin y Yusuf	1995	Bangladesh	S, Q	Y	CS	DS
Neill, Davalos, Kiiru y Sebstad	1994	Kenya	S	N	CS	
Nelson	1984	Indonesia	S	Y	CS	
Nelson y Bolnick	1986	Indonesia	S	Y	CS	
Oldham, Hadidid, Hussein, Aziz y Sakr	1994	Egipto	S	N	CS	
Park y Ren	2001	China	S	Y	CS	
Pitt y Khandker	1996	Bangladesh	S	Y	CS	SB, DS
Pitt y Khandker	1998	Bangladesh	S	Y	CS	SB, DS
Pitt, Khandker, Chowshury y Millimet	1998	Bangladesh	S	Y	CS	SB, DS
Pitt, Khandker, McKernan y Latif	1999	Bangladesh	S	Y	CS	SB, DS, CL
Pulley	1989	India	S	Y	L	
Schuler y Hashemi	1994	Bangladesh	S, Q	Y	L	SB, DS, CL
Shuler, Hashemi y Badal	1998	Bangladesh	Q	Y	L	
Schuler, Hashemi y Riley	1997	Bangladesh	S, Q	Y	L	SB, DS, CL
Sebstad	1992	Sudáfrica	Q	N	CS	
		Bangladesh	S, Q	Y	CS	
Sebstad y Cohen	2002	Bolivia	S, Q	Y	CS	
		Filipinas	Q	N	CS	
		Uganda	Q	N	CS	
Sebstad y Loza	1993	Egipto	S, Q	N	CS	
Sebstad y Walsh	1991	Kenya	S, Q	N	CS	
Smith	2002	Ecuador	S	Y	L	
		Honduras	S	Y	L	
Steele, Amin y Naved	2001	Bangladesh	S	Y	L	
Sutoro	1990	Indonesia	S, Q	N	CS	
Todd	1996	Bangladesh	Q	Y	CS	
Vengroff y Creevey	1994	Senegal	S, Q	Y	CS	
Woller y Parsons	2002	Ecuador	S	N	CS	CL
Wydick	1999a	Guatemala	S	Y	CS	
Wydick	1999b	Guatemala	S	Y	CS	
Zaman	2001	Bangladesh	S	Y	CS	SB, F

Tabla 17.1 Resumen de EI reseñadas y analizadas (cont.)

Evaluación de impacto reseñada ^a	Año	País	Metodología ^b	Grupo de control ^c	Marco temporal ^d	Comentario ^e
Zeller, Ahmed, Babu, Broca, Diagne y Sharma	1996	Camerún	S	Y	L	CL
		Malí	S	Y	L	CL
		Ghana	S	Y	CS	CL
		Nepal	S	Y	L	CL
		Pakistán	S	Y	L	CL
		China	S	Y	CS	CL
		Bangladesh	S	Y	L	CL
		Madagascar Malawi	S S	Y Y	L L	CL CL
Países de la OCDE						
Ashe y MacIntyre	2002	E.U.A.	S, Q	N	CS	
Benus, Word y Grover	1994	E.U.A.	S	Y	L	SB
Blair y Klein	2001	E.U.A.	S, Q	N	L	
Clark y Huston	1993	E.U.A.	S, Q	N	L	
Clark, Kays, Zandniapour, Soto, Doyle	1999	E.U.A.	S, Q	N	L	
Drury, Walsh y Strong	1994	E.U.A.	S	N	L	
Dumas	2001	E.U.A.	Q	N	CS	
Else y Clay-Thompson	1998	E.U.A.	Q	N	CS	BC
Himes y Servon	1998	E.U.A.	S, Q	N	CS	DS
Kosanovich y Fleco	2002	E.U.A.	S	Y	L	BC
Instituto para el Desarrollo Social y Económico	1994	E.U.A.	S	Y	L	
Mt. Auburn Associates	1998	E.U.A.	S	N	CS	DS
Raheim	1996	E.U.A.	S	N	CS	
Raheim y Friedman	1999	E.U.A.	S, Q	N	L	
Sekkesaeter	2002	Noruega	S, Q	N	CS	
The Roberts Foundation	1995	E.U.A.	S	N	CS	
Departamento de los E.U.A. para la Salud y los Servicios Humanos	1994a	E.U.A.	S	Y	L	
Departamento de los E.U.A. para la Salud y los Servicios Humanos	1994b	E.U.A.	S	Y	L	BC
Departamento de los E.U.A. para la Salud y los Servicios Humanos	1994c	E.U.A.	S	Y	L	SB
Departamento de los E.U.A. para la Salud y los Servicios Humanos	1994d	E.U.A.	S, Q	Y	CS	

a) Diversas evaluaciones del impacto reseñadas se basaron por completo o en parte en la misma evaluación o datos de evaluación del programa. Estos incluyen Coleman (1999, 2001b); Nelson (1984), Nelson y Bolnick (1986), Bolnick y Nelson (1990); Khandker (1994), Pitt y Khandker (1994, 1998), McKernan (1996), Pitt, Khandker, Chowdhury, Millimet (1998), Khandker, Samad y Khan (1998), Pitt, Khandker, McKernan y Latif (1999), y Khandker (2001); Schuler y Hashemi (1994), Hashemi y Riley (1996), Schuler, Hashemi y Riley (1997), y Schuler, Hashemi y Badal (1998); Wydick (1999a, 1999b) y Kevane y Wydick (2001).

b) S= encuesta de impacto; Q= evaluación cualitativa.

c) Y= sí; N = no.

d) L= longitudinal; CS= corte transversal.

e) SB= controles para el sesgo de selección por características no observables; DS= administra encuestas en diferentes épocas/temporadas al año; CL= Evalúa los beneficios comunitarios; F= controles para fungibilidad de los préstamos; BC= realiza un análisis costo-beneficio.

f) No es una evaluación de programa, sino un estudio de empresas rurales que no son granjas.

Lo mismo es poco cierto en los países de la OCDE. En relación con los LDCs, los países de la OCDE poseen mercados de servicios financieros altamente desarrollados a los cuales los pobres tienen acceso (si acaso de manera limitada en ocasiones). Además, hay relativamente poca discusión entre los defensores de la microempresa de la penetración financiera; en su lugar, los objetivos de la industria tienden a estar definidos dentro del contexto del autoempleo como una alternativa al sector formal de empleo. Otra diferencia importante entre microfinanzas y microempresa es que la última enfatiza en mayor grado la capacitación empresarial. La integración de los préstamos con el entrenamiento empresarial es común dentro de la industria microempresarial, pero la capacitación es rara y es objeto de muchas disputas dentro de la industria de las microfinanzas.

Por lo tanto, debido a las diferencias que existen entre ambas, aquí se tiene el cuidado de diferenciar entre ellas donde es relevante. Dicho lo anterior, los aspectos expuestos en este estudio se aplican más o menos de la misma forma a la evaluación de impacto de ambos tipos de programas. Además, la necesidad de prácticas sanas para la evaluación del impacto se aplica por igual a ambas, así como lo hacen los principios metodológicos para llevarlas a la práctica.

Tabla 17.2 **Resumen de estudios de la EI**

	LDC	OCDE	Total
Estudios de impacto publicados	67	20	87 ^a
Evaluaciones del impacto realizadas	90	20	110 ^b
Metodología			
Sólo encuesta	58	10	68
Encuesta + Cualitativo	26	8	34
Sólo cualitativo	6	2	8
Grupo de control			
Sí	73	7	81
No	17 ^c	13	28
Marco temporal			
Longitudinal	27	11	38
Transversal	63	9	72
Control para el sesgo de selección	19	2	21
Aplicación de encuestas en diferentes temporadas	14	0	14
Evaluación de los impactos a nivel comunitario	24	0	24
Control por la sustitución de los recursos de los apoyos	2	0	2
Efectúan análisis costo-beneficio	6	2	8
Incluyen los abandonos del programa	1	2	3
Número de países	31	2	33

- Se refiere a estudios de la EI publicados, sin importar el número de evaluaciones del impacto independientes cubiertas en cada estudio.
- Se refiere al número de evaluaciones del impacto encontradas en cada estudio de la EI. Evaluaciones del impacto de dos o más programas en el mismo país, como parte de una evaluación integrada del impacto, fueron contadas como una sola evaluación. Las evaluaciones de impacto de programas en diferentes países pero publicadas en el mismo estudio de la EI fueron contadas como evaluaciones diferentes.
- Los grupos de control no fueron relevantes para la metodología para el impacto económico comunitario utilizada por Woller y Parsons (2002).

Fundamentos conceptuales de la evaluación del impacto

El impacto ocurre teóricamente en cuatro ámbitos: el individual, la empresa, el hogar y la comunidad.³ En teoría, la cadena causal del impacto funciona de una manera parecida a la siguiente: 1) los préstamos y la capacitación llevan a un incremento en la formación y expansión empresarial y a una mayor inversión en capital de trabajo y en activos productivos; 2) la mayor formación, expansión e inversión empresarial lleva a un incremento en los retornos empresariales; 3) los mayores retornos empresariales llevan a una mayor creación de trabajo y a un mayor ingreso para el hogar; 4) el mayor ingreso del hogar lleva a mayores niveles de consumo del hogar, acumulación de activos, inversión en el recurso humano e inversión en activos físicos. A su vez, el mayor ingreso del hogar y la acumulación de activos, en conjunto con un mayor nivel de acceso a los servicios financieros, expanden las estrategias *ex ante* y *ex post* de reacción y supervivencia de los hogares pobres, por lo tanto los hace menos vulnerables al riesgo.

Tabla 17.3 Países en los cuales se efectuó la evaluación de impacto

	#
Bangladesh	22
Bolivia	5
Brasil	1
Camerún	1
China	2
República Dominicana	1
Ecuador	3
Egipto	2
Ghana	2
Guatemala	3
Guinea	1
Honduras	2
India	4
Indonesia	6
Jamaica	1
Kenia	4
Madagascar	1
Malawi	4
Malí	1
Nepal	2
Noruega	1
Pakistán	1
Perú	3
Filipinas	3
Senegal	1
Sudáfrica	2
Sri Lanka	2
Tailandia	3
Uganda	2
Estados Unidos	20
Alto Volta	1
Zambia	1
Zimbabwe	1

Bajando al nivel individual, el acceso a los servicios financieros, el control sobre los recursos financieros, la propiedad y la operación empresarial, la mayor contribución del ingreso del hogar, las redes de grupo y el apoyo mutuo llevan a mayores niveles de facultamiento personal y social, especialmente entre las participantes femeninas del programa. En el ámbito comunitario, los beneficios creados a los otros tres niveles crean externalidades positivas que se difunden a través de las comunidades locales y aldeañas. (La Tabla 17.4 lista indicadores comunes utilizados para medir los impactos en cada uno de los cuatro niveles de impacto).

Tabla 17.4 Indicadores de impacto en los niveles individual, empresarial, del hogar y comunitario

Nivel de impacto	Indicadores
Individual	<ul style="list-style-type: none"> • Toma de decisiones intra-hogar (participación en la toma de decisiones del hogar sobre aspectos como las finanzas, la escuela, la salud, la planeación familiar, etc.) • Control sobre recursos financieros y de otro tipo • Contribución a los ingresos del hogar • Uso de anticonceptivos • Autoestima • Actitudes ante uno mismo, la vida y el futuro • Conciencia política y social • Participación en las esferas sociales y políticas • Abuso de la pareja
Empresarial	<ul style="list-style-type: none"> • Ventas • Ganancias • Valor neto • Propiedad y adquisición de activos • Empleos creados • Diversificación de productos • Diversificación del negocio • Adopción de prácticas de negocio
Del Hogar	<ul style="list-style-type: none"> • Ingreso <ul style="list-style-type: none"> – Gasto – Gasto en alimentos – Gasto en tipos específicos de alimentos (<i>v. gr.</i> frutas, vegetales, carnes, lácteos) – Gasto en medicamentos y la salud – Gasto en la educación de los hijos • Propiedad y adquisición de activos • Ahorros • Inversión en vivienda y reparación del hogar • Inversión en terrenos • Acceso y uso de medicinas y servicios de salud • Conocimiento y uso de prácticas simples de higiene • Conocimiento y uso de intervenciones médicas/prácticas de salud simples (<i>v. gr.</i> terapia de rehidratación oral, alimentación de pecho) • Asistencia de los hijos a la escuela • Tipos y frecuencia de los alimentos consumidos • Incidencia y duración de las "temporadas de hambre" • Medidas antropomórficas de los hijos • Respuesta a e impacto de los choques económicos y de otro tipo

Tabla 17.4 **Indicadores de impacto en los niveles individual, empresarial, del hogar y comunitario** (cont.)

Nivel de impacto	Indicadores
Comunitario	<ul style="list-style-type: none"> • Asistencia infantil a la escuela • Uso de anticonceptivos • Empleos creados/tasa de empleo • Ingreso • Gasto • Valor neto • Producción • Salarios • Precios • Participación en las esferas políticas y sociales • Participación en apoyo de las familias • Pobreza

En la práctica, la cadena causal es más compleja que la descrita antes. Al menos, una multiplicidad de factores intermedios influyen el impacto. Los ejemplos incluyen los atributos del programa, las características de los clientes, la geografía, la estructura social y las relaciones de poder, la infraestructura física y económica, y la macroeconomía. Algo más son las relaciones recíprocas entre causas y efectos, en las cuales los impactos se vuelven causas, las causas se vuelven impactos, los impactos, causas nuevamente, y continúa de esta forma, de manera que es crecientemente difícil distinguir entre unas y otros. Finalmente, la sustitución de los recursos de los apoyos hace el rastreo, a través de la secuencia exacta de causa y efecto, virtualmente imposible (véase la discusión más adelante), creando así algo más parecido a una red causal que a una cadena causal. El resultado final es que los individuos experimentan los impactos de maneras diferentes y que no existen dos programas que creen impactos iguales de la misma manera. Así que quizás no es sorprendente que la única consistencia en los hallazgos de la EI es su inconsistencia —una evaluación del impacto típica produce hallazgos mixtos, y los hallazgos varían considerablemente de estudio a estudio—. Esta inconsistencia subraya los peligros inherentes de asumir una relación causal entre la participación en un programa y cualquier resultado de política específico.

Enfoques metodológicos de la evaluación del impacto

Con el marco conceptual a la mano, los investigadores de la EI pueden seleccionar entre diversos enfoques metodológicos para llevar a cabo las evaluaciones del impacto. Dependiendo del propósito, los enfoques pueden ubicarse en uno o dos paradigmas metodológicos (o ambos): el paradigma de “prueba” y el paradigma de “mejora”. En el paradigma de prueba, el propósito de la EI es atribuir la causalidad de los resultados observados a la participación del programa. El paradigma de prueba adopta el lenguaje, metodología y visión

del mundo de las ciencias físicas y sociales. Su audiencia es principalmente externa —donantes, diseñadores de política, académicos— de los cuales el rigor metodológico y la validez científica son las principales virtudes.

Dentro del paradigma de mejora, el objetivo de una EI es mejorar el impacto de los servicios financieros en los participantes de los programas mediante la mejora en los productos y políticas. El paradigma de mejora adopta el lenguaje, metodología y visión del mundo de la administración. Su audiencia es principalmente interna —administradores, equipo y clientes— para quienes la utilidad, reducción del tiempo y el costo son consideraciones fundamentales.

Dentro de este contexto de paradigmas de prueba y de mejora, se han identificado cinco enfoques metodológicos:

1. El método científico, el cual se basa en las ciencias naturales.
2. La tradición humanista, la cual utiliza la teoría y corrobora la evidencia para hacer juicios razonados.
3. Evaluaciones de alcance medio, en las cuales se toma en cuenta explícitamente las restricciones impuestas por “el campo de la realidad”.
4. Aprendizaje participativo, en el cual facilita la articulación subjetiva de los participantes “la realidad” para llegar a conclusiones informadas.⁴
5. Investigación de mercado, la cual enfatiza la recopilación y uso de la inteligencia de los mercados para informar a los administradores en la toma de decisiones.

Si nosotros asumimos un continuo con el paradigma de prueba en un extremo y el paradigma de mejora en el otro, el método científico y la investigación de mercados serían representantes de los respectivos extremos. Los tres métodos faltantes tenderían a juntarse al centro del continuo. El método científico se basa fundamentalmente en evidencia cuantitativa y la tradición humanista y el aprendizaje participativo en la cualitativa. Las evaluaciones de alcance medio y de investigación de mercados tienden a basarse en diversas combinaciones de evidencia cualitativa y cuantitativa. En la práctica, estos cinco enfoques pueden ser utilizados tanto para probar como para mejorar. Además, es común que los investigadores utilicen una combinación de enfoques, junto con combinaciones de métodos cuantitativos y cualitativos, para poder comprobar los datos y añadir profundidad, confiabilidad y relevancia a sus hallazgos.

Método científico

En un experimento científico clásico, los sujetos de estudio se retoman de la misma población, comparten características comunes, y son seleccionadas aleatoriamente tanto en el grupo de tratamiento (aquellos que reciben la

intervención) como en el grupo de control (aquellos que no reciben la intervención). El propósito de un experimento clásico es controlar todos los factores del medio para tener la capacidad de atribuir cualquier diferencia observada en los resultados obtenidos entre el grupo de tratamiento y el de control debido a una intervención. Desafortunadamente, en las ciencias sociales las condiciones para un experimento clásico raras veces existen, y la capacidad del investigador para controlar todos los factores del medio es ampliamente limitada, si no es que imposible.

Una breve dimensión del experimento clásico es el “quasi-experimento” el cual tiende a replicar lo más posible, las condiciones del experimento clásico dentro de las restricciones existentes y generalmente utiliza instrumentos de investigación basados en encuestas. Otra forma de pensar en quasi-experimentos es en la prueba “con-sin” en la cual el investigador debe establecer el contrafactual de lo que hubiera sucedido si el grupo de tratamiento no hubiese recibido la intervención.

La regresión múltiple es una forma de quasi-experimento. Prueba el impacto de ciertas variables explicativas en los resultados observados, al mismo tiempo que mantiene otros factores del medio constantes. La regresión múltiple no es ampliamente utilizada en evaluaciones de impacto, sin embargo, ante grandes demandas de datos es necesario tomarla en cuenta para todos los factores del medio relevantes, la experiencia técnica requerida para representarlas e interpretarlas, y lo poco claro de los supuestos econométricos necesarios para validarlas.

Los quasi-experimentos más frecuentes son los estudios de grupo de control. Como su nombre lo indica, los estudios de grupo de control emparejan los resultados en un grupo de tratamiento de un programa de participantes con los de un grupo de control de no participantes. Si todo va bien, el grupo de control encaja con el grupo de tratamiento en características clave, y el investigador puede atribuir razonablemente las diferencias observadas en los resultados a la participación en el programa. Sin embargo, no siempre va todo bien. Los estudios de grupo de control se enfrentan a dificultades metodológicas potenciales. Sucumbir ante dichas dificultades puede comprometer seriamente la validez del estudio.

Siguen, en riguroso orden, las evaluaciones de impacto que aspiran al método científico, pero que están más propiamente caracterizadas como pseudo-científicas, debido a que violan ciertos principios científicos clave. Un ejemplo común de éstas son las evaluaciones “antes-después”, las cuales comparan los resultados entre los participantes del programa en un punto (v.g., después de unirse al programa) con otro punto anterior (v.g., antes de unirse al programa), pero no utilizan grupos de control. La ausencia de grupos de control en la

evaluación “antes-después” hace estadísticamente inválido atribuir el impacto a la participación en el programa.

Tradición humanista

La tradición humanista es un enfoque inductivo de aprendizaje que conjunta una amplia gama de herramientas desarrolladas y refinadas por los científicos sociales en áreas como la sociología y la antropología. Tomando prestada la libertad de la etnografía, el enfoque humanista involucra el estudio de pequeños grupos de sujetos dentro de su propio ambiente. En lugar de buscar pequeños grupos de variables y un gran número de casos, busca obtener una comprensión detallada de las circunstancias sobre los pocos sujetos a ser estudiados. Es descriptivo e interpretativo, descriptivo debido a que el detalle es crucial e interpretativo debido a que el investigador debe determinar el significado de lo que él o ella observa sin reunir amplias pruebas o información estadística. Los métodos incluyen informantes clave, entrevistas de profundidad con participantes, estudios de caso, y observación de participantes (*v.g.*, prolongación de la residencia de investigadores de campo en comunidades que reciben el programa).

La tradición humanista no busca probar el impacto en ningún sentido estadístico, pero ofrece una interpretación de la relación entre la participación en el programa y los resultados que se obtuvieron con un alto grado de verosimilitud, lo cual permite la inferencia y causabilidad entre los resultados observados y la participación en el programa. La verosimilitud depende de factores tales como la evidencia, la metodología de la muestra, la consistencia lógica, la calidad de la evidencia y el razonamiento, el grado de confirmación de la evidencia vía triangulación y fuentes secundarias de información, y la reputación del investigador.

Con relación al método científico, la tradición humanista carece de un grupo bien definido de valores metodológicos. Cómo dichos valores podrían especificarse para guiar a los investigadores *ex ante* y ayudar a los diseñadores de políticas a indicar la calidad del trabajo *ex post*, es aún una pregunta sin respuesta. Sin embargo, la experiencia ha demostrado que este enfoque puede conseguir resultados con una razonablemente alta confiabilidad y pueden proveer hallazgos amplios y profundos que no siempre se pueden obtener a través de los métodos científicos. Pueden producir, a su vez, conclusiones de aún más valor para el método científico, que particularmente el caso de “evaluaciones de impacto basadas en encuestas que se hacen pasar por ciencia, pero que no recolectan datos con rigor científico” (Hulme, 2000, p. 87).

Aprendizaje Participativo y Acción (PLA, por sus siglas en inglés)

El Aprendizaje Participativo y Acción comprende una gran variedad de metodologías, incluyendo la Estimación Rural Participativa (PRA), Estimación Rural Rápida (RRA), Métodos de Aprendizaje Participativo (PALM), Investigación de Acción Participativa (PAR). PLA asume la existencia de múltiples realidades subjetivas. Requiere primero que el investigador responda a la pregunta “¿La realidad de quien cuenta? (Chamers 1997). El trabajo del investigador es después elucidar sobre la realidad compartida por el grupo objetivo a través del proceso de “creación del conocimiento”, definido como la participación completa del grupo objetivo en la identificación, análisis y plan de acción para resolver el problema. Los métodos del PLA incluyen actividades tales como visualización, calendarios estacionales, horizontes temporales históricos, diagramas de Venn, revisiones transectoriales, discusiones en grupos de enfoque, catalogación de preferencias relativas y entrevistas semiestructuradas.

PLA ofrece el más radical y serio desafío al método científico al aplicarlo al desarrollo internacional. El grupo de defensores del PLA no se centra en escuchar lo que ellos ven como los defectos del método científico (Hulme, 2000):

- Ignoran la complejidad, diversidad y contingencia de los hogares pobres, su sustento y estrategias de supervivencia.
- Conciben a la causalidad como una cadena simple, unidireccional y no como la compleja red que es.
- Miden lo trivial o pretenden medir lo que no puede ser medido.
- Es extractivo y explotador.
- Refuerza el *status quo* a través de otorgar poderes a tecnócratas, expertos, profesionistas, diseñadores de política y élites.
- No guía hacia la acción propositiva para y en nombre de los grupos más pobres.

Aunque ciertas, esas son críticas mordaces. Si esos alegatos son ciertos es quizá una cuestión de la propia realidad subjetiva, aunque quizá probablemente contienen algo de verdad. Cuando menos, requieren de cierto nivel de humildad y autoreflexión crítica entre los profesionistas del método científico, dos rasgos que no siempre son evidentes.

PLA tiene su propio conjunto de debilidades. Entre ellas están las inherentes a la subjetividad, la falta de estandarización (lo cual dificulta las comparaciones) su pluralismo (lo cual genera perspectivas conflictivas sobre el impacto), su ingenuo supuesto de que la participación equilibra a la representación dadas las relaciones de poder locales, su afirmación de que la participación es equivalente al facultamiento, y su falta de transparencia (lo cual dificulta las evaluaciones *ex post* con rigor metodológico).

Una cuestión final es la atribución. Como la tradición humanista, PLA tampoco puede alegar la causalidad en términos científicos. Se basa en una triangulación de evidencias, profundidad del conocimiento, calidad del método y de la información, y la habilidad del investigador para establecer un vínculo de causalidad posible. Dichas limitaciones, sin embargo, no impiden a sus defensores argumentar que estudios participativos bien dirigidos pueden generar resultados más fiables que las encuestas convencionales (ver, por ejemplo, Chambers, 1997, pp. 141-146.)

Como en la tradición humanista, la calidad y fiabilidad de los estudios PLA, varía ampliamente, dependiendo de factores tales como las habilidades de los facilitadores, la motivación de los grupos objetivo, la aplicación de las herramientas a determinadas situaciones, o el grado de participación.⁵ En la práctica, los defensores del PLA han sido muy pragmáticos. Son indecisos al prescribir las mejores prácticas específicas, prefiriendo dejar a los profesionistas los mejores juicios para adaptar la metodología y las herramientas a las circunstancias.

Evaluaciones de alcance medio

Las evaluaciones de impacto de alcance medio provienen de la percepción extendida entre los profesionistas de las micro finanzas de que el método científico está desconectado tanto de las realidades en campo como de las necesidades del administrador. Los profesionistas se han quejado de que las evaluaciones de impacto basadas en encuestas son costosas, largas, agobiantes, e inoportunas, requieren experiencia técnica y recursos que van más allá de las capacidades institucionales; y están enfocadas a audiencias externas, con poca atención a la utilidad administrativa. Ni los organismos o fundaciones o los evaluadores externos invierten tiempo y dinero en construir y fortalecer la capacidad institucional para hacer una evaluación de impacto. Agregan a eso la presión de los donadores de adquirir la autosuficiencia institucional y la evaluación de impacto es vista ante todo como un tema más del reporte de gastos. La falta de recursos, de habilidades técnicas y de material incentiva a que la mayoría de los programas no cuenten con evaluaciones de impacto, el resultado es que en general no se tenga idea de cuál ha sido el impacto del programa.

Las evaluaciones de alcance medio son producto de un esfuerzo de profesionistas coordinados por el proyecto AIMS (ver nota al pie 2) para corregir las deficiencias del método científico, vincular a los paradigmas de prueba y de mejora y construir la capacidad institucional para la evaluación de impacto. El resultado final de este esfuerzo fue la creación de un grupo de herramientas “profesionistas-amigables” de evaluación de impacto diseñadas para tomar en consideración las realidades de campo, generar resultados útiles para la gestión,

y para que pueda ser implementada por el *staff* del programa. Las herramientas SEEP/AIMS, como han sido denominadas,⁶ incorporan metodologías de cada uno de los cuatro enfoques metodológicos de evaluación de impacto. Constan de dos herramientas cuantitativas y tres cualitativas: la encuesta de impacto, la encuesta de salida del cliente, grupos de enfoque de satisfacción del cliente, entrevistas a profundidad sobre facultamiento y entrevistas temporales sobre ahorro y crédito.⁷ Las evaluaciones de alcance medio no aspiran a probar el impacto pero en su lugar buscan establecer posibles asociaciones entre los resultados observados y la participación en el programa.

Las herramientas SEEP/AIMS han disfrutado de una legitimidad respetable entre las organizaciones de profesionistas desde su introducción. A la fecha, docenas de instituciones de microfinanciamiento han recibido capacitación formal e intensiva en el uso de las herramientas, y muchas han sido exitosas al implementarlas en sus respectivos campos de trabajo. Las evaluaciones de alcance medio sin embargo, no se limitan a las herramientas SEEP/AIMS, ni tampoco todos los profesionistas están satisfechos con ellas. El trabajo continúa para refinar dichas herramientas y adaptarlas a los contextos locales, o bien para desarrollar herramientas aún más “amigables para los profesionistas” (*v. gr.*, “AIMS-Lite”).

En su mayor parte, las evaluaciones de alcance medio se adhieren y toman prestados estándares y enfoques metodológicos. La principal concesión que hacen es recomendar el uso del personal del programa para conducir la investigación y, relacionado con el impacto de la encuesta, para dirigir las evaluaciones a través de muestras representativas transversales (en oposición a longitudinales) utilizando los llamados clientes “pivote” —clientes reclutados a través de las operaciones normales del programa pero que aún no han recibido los préstamos como grupo de control. (Véase en el apéndice la lista de recomendaciones para dirigir evaluaciones de alcance medio).

Investigación de mercado

Integrado al paradigma de mejora, la evaluación de impacto y la investigación de mercado están ampliamente entrelazadas: el conocimiento oportuno acerca del impacto le dice al administrador del programa qué tan efectivos son sus programas y políticas, el conocimiento del mercado por lo tanto permite al administrador del programa diseñar productos y políticas que mejoren el impacto. La investigación de mercado (definida como la recopilación, análisis y uso de la inteligencia de los mercados) juega por tanto un papel central dentro del paradigma de mejora. El surgimiento de la investigación de mercado en la agenda de micro finanzas se da debido a diversas tendencias del mercado: 1) la competencia en la industria está aumentando y se espera que aumente aún más; 2) Los consumidores de micro finanzas se están convirtiendo en más

concededores, exigentes y asertivos; 3) los clientes están abandonando programas de micro finanzas a tasas alarmantes, y, en consecuencia, 4) las instituciones de micro finanzas están adoptando estrategias y prácticas comerciales.

La investigación de mercado usa una gran variedad de herramientas cualitativas y cuantitativas, incluyendo encuestas, discusiones en grupos de enfoque, entrevistas a profundidad y evaluaciones participativas.⁸ La investigación de mercado incluye las actividades ocasionales de evaluación y como pasa de la información del cliente y del mercado a programas operacionales y sistemas de información administrativos. La investigación de mercado hace muy poca referencia a la validez científica. La falta de atribuciones es una debilidad particular, principalmente dadas las dificultades logísticas para integrar a grupos de control de no clientes en los sistemas rutinarios de obtención de datos. Desde la perspectiva administrativa, sin embargo, la pérdida de certeza estadística es mayor a la ganancia en la velocidad, costo y utilidad. Dejando atrás la herramienta de desarrollo en la investigación de mercados, se encuentra el desarrollo y establecimiento de un conjunto de indicadores metodológicos y directrices para la implementación en campo de las herramientas de la investigación de mercados. (para mayor información sobre la relación entre evaluación de impacto e investigación de mercados, véase Cohen 1999 y Copestake 2000.)

Entre los cinco enfoques revisados arriba, el método científico y las evaluaciones de alcance medio tienen los indicadores más claros y articulados para conducir una metodología *ex ante* y para evaluar la calidad de la metodología *ex post* (dado que la aspiración de las evaluaciones de alcance medio es la verosimilitud científica, éstas son tema para muchos de los mismos indicadores metodológicos del método científico). Más aún, los métodos científicos basados en encuestas dominan los estudios de las evaluaciones de impacto. En la mayoría de los estudios los métodos cuantitativos (ya sea que se encuentren en la tradición humanista o en la PLA) son utilizados como suplemento de la investigación científica basada en encuestas. De las evaluaciones de impacto revisadas, 68 utilizaban encuestas como el único instrumento de investigación y 34 utilizaban encuestas como principal instrumento de investigación, complementándolo con métodos cualitativos. Sólo 8 de las evaluaciones revisadas se basaban únicamente en métodos cualitativos, incluyendo Goetz y Gupta (1996), Todd (1996), Schuler *et al.* (1997), Schuler *et al.* (1998), Dumas (2001), y las evaluaciones de programas en Filipinas y Uganda resumidas en Cohen y Sebstad (2000). A la luz del dominio de la evaluación de impacto científica y sus relativamente claros indicadores metodológicos, el resto de este estudio se enfoca en temas metodológicos relevantes para la evaluación de impacto científica.

Obstáculos metodológicos de la evaluación de impacto científica

Como mencioné anteriormente, las condiciones ideales para dirigir una evaluación de impacto científica raras veces existe. Los investigadores de la evaluación de impacto deben frecuentemente elegir una segunda o peor opción. Al enfrentarse a las restricciones siempre presentes, los investigadores tienen que preguntarse a sí mismos hasta qué grado ellos están dispuestos a comprometer los principios metodológicos aceptados para adaptarse a esas restricciones. En otras palabras, “¿Qué costo en términos de precisión científica están dispuestos a aceptar a cambio de obtener la viabilidad en la implementación?”

En respuesta, el aspecto acerca de intercambios razonables es útil para conocer cuáles son los principales obstáculos al conducir evaluaciones de impacto científicas. Esta sección revisa dichos obstáculos. Por orden de presentación, son: 1) construcción de grupos de control válidos, 2) el sesgo de selección de la muestra a partir de características individuales observables, e inobservables de los individuos, fallas para considerar a las personas que abandonaron el programa y la ubicación no aleatoria del programa; 3) contaminación en el grupo de control; 4) tendencias frecuentes; 5) viabilidad de los préstamos; y 6) horizonte temporal de la evaluación de impacto (estudios longitudinales y transversales).

Grupos de control válidos

Algunas evaluaciones de impacto científicas que buscan inferir la causalidad o la posible asociación entre resultados y la participación en el programa, requieren de grupos de comparación para validar a los grupos de control de no clientes. Construir un grupo de control válido, sin embargo, puede ser difícil. Hay diversas razones por las cuales la construcción de un grupo de control válido puede no ser factible, entre ellos existe una fuente de vinculación o restricciones del ambiente (temas comunes para los profesionistas encargados de las evaluaciones de impacto) y, como se explica abajo, los retos para la construcción y rastreo el grupo de control que suponen las evaluaciones de impacto longitudinales (series de tiempo).

Si a causa de las limitaciones de carácter vinculante u otras razones, varias de las evaluaciones de impacto que se han revisado para este estudio no utilizan grupos de control. La ausencia de grupos de control es particularmente notoria en la evaluación de programas microempresariales; sólo 7 de 20 evaluaciones de programas microempresariales utilizaron un grupo de control en comparación con 73 de 90 evaluaciones de programas de micro financiamiento.

La incapacidad para inferir causalidad no significa que las evaluaciones de impacto que carecen de grupos de control no tengan valor, pero sí que su valor radica en otra parte. Ellos pueden, por ejemplo, ser herramientas valiosas para monitorear el progreso de los clientes, evaluar resultados relativos entre

diferentes sectores del mercado, medir los resultados frente los objetivos del programa, o calcular el costo unitario para resultados específicos del programa, los cuales son información útil para los administradores, donantes y diseñadores de política. No son útiles, sin embargo, para determinar si la participación en el programa está causalmente relacionada con los resultados de política deseada, tales como reducción de la pobreza o creación de empleo.

Básicamente, la selección del grupo de control requiere de identificar al grupo de personas que no están participando en un programa de capacitación o de crédito y que comparten características similares al grupo de tratamiento y aleatoriamente seleccionar de entre ellos. No todos los que pertenezcan a la población de no clientes, sin embargo, son candidatos de un grupo de control legítimo. La construcción de un grupo de control válido requiere que el grupo de control se empareje con el grupo de tratamiento en las características clave observables y no observables. Las fallas para realizar eso crea el llamado “sesgo de selección”, el cual es el principal obstáculo metodológico que aqueja a los estudios de grupo de control. Dependiendo de su seriedad, al sesgo de selección se le atribuyen los impactos observados en cualquier lugar problemático e invalida totalmente la muestra. Veinte de 90 evaluaciones en países poco desarrollados buscaban controlar a través de sesgar la selección de la muestra, comparado con dos de las 20 evaluaciones de programas de la OCDE.

Sesgo de selección

El sesgo de selección proviene de cuatro fuentes principales: 1) fallas para emparejar las características observables en los individuos de un grupo de tratamiento y uno de control, 2) fallas para emparejar las características no observables en los individuos de un grupo de tratamiento y de uno de control, 3) fallas para tomar en cuenta a los que se salen del programa y 4) programas de colocación no aleatorios. Éstas serán consideradas posteriormente.

Fallas para emparejar las características observables de los individuos

Las diferencias de los resultados entre el grupo de tratamiento y el de control puede darse en gran medida, en función de las diferencias en las características observables, tales como edad, género, educación, situación de autoempleo, tipo de empresa, o ubicación geográfica, que debido a la participación en el programa. La construcción de un grupo de control válido requiere entonces que los miembros del grupo compartan características observables similares. Por ejemplo, si el perfil del grupo de tratamiento es 85 por ciento femenino, edad de 25-50, de 0-5 años de educación formal, autoempleados, rurales, y más o menos distribuidos entre actividades de manufactura, minorista y de servicios, un grupo de control válido deberá emparejar estas características observables y acercarse lo más posible.

Emparejar a los grupos de tratamiento y de control con las características observadas puede ser retador, pero eso no significa que sea insuperable. Un buen investigador debe ser capaz de evitar este obstáculo con comparativa facilidad. Para asegurarse de que el grupo de control coincide de cerca con el grupo de tratamiento, la muestra aleatoria estratificada con grupos seleccionados específicamente entre la población objetivo, puede ser necesaria. Si el grupo de tratamiento y el de control coinciden razonablemente en las características observables podrá ser fácilmente determinado comparando las medias de los grupos en las características clave observadas y al probar que si hay diferencias en las medias son estadísticamente significativas.

Fallas para emparejar las características inobservables de los individuos

Una fuente de error para sesgar la selección que aqueja aún más a los investigadores es fallar al emparejar al grupo de control y de tratamiento en características inobservables de los individuos lo cual también podría tener un impacto en los resultados. Uno podría por ejemplo preguntarse, ¿por qué una persona se incorpora a un programa de microfinanciamiento y otras no? o ¿por qué unas personas se incorporan rápidamente y otras tardan más en incorporarse? Las respuestas a esas preguntas pueden ser probablemente multifacéticas, pero una hipótesis razonable es que al final, los que se incorporan antes o después poseen ciertas características no observadas que no tienen los que no se incorporan, como podrían ser la conducción empresarial, el deseo de asumir el riesgo, un ambiente de apoyo en el hogar, o simplemente, la determinación de mejorar su propia vida. La respuesta puede también reflejar beneficios netos esperados de la participación en el programa. Los que se incorporan rápidamente pueden tener un mayor beneficio neto esperado *ex ante* de su participación que los que no se incorporan o que tardan más en incorporarse. Ellos tenderán, como resultado, a disfrutar de mayores beneficios netos *ex post*. El punto es que las características no observables de los individuos pueden posiblemente ser clave al determinar los impactos de la participación en el programa.

El mejor método para controlar el sesgo en la selección de la muestra para las características no observables es a través del muestreo aleatorio de participantes del grupo de control y del de tratamiento. La selección aleatoria entre los estudios de evaluación de impacto es, sin embargo, rara, lo que sugiere que los estudios de evaluación de impacto generalmente exageran los efectos e impactos de los programas. Entre los pocos estudios de evaluación de impacto que utilizan métodos de asignación aleatoria se encuentran: Benus *et al.* (1994), US Department of Health and Human Services (1994c), MKNelly y Dunford (1999a, 1999b), y Coleman (1999b, 2001). Un buen ejemplo de selección aleatoria está en las demostraciones de la evaluación del autoempleo en los estados de Washington

y Massachussets. Los investigadores invitaron a los solicitantes desempleados interesados en auto emplearse a asistir a una sesión informativa que daba una explicación básica sobre los riesgos y beneficios del autoempleo y las principales características del programa. Al final de esa sesión, los solicitantes que aún estaban interesados recibían una cédula para aplicar al programa. Aquellos que completaban la cédula a tiempo y que cumplían con los requisitos de elegibilidad eran aleatoriamente seleccionados para ser grupo de tratamiento elegible para recibir los servicios para el desarrollo empresarial y asistencia financiera o a un grupo de control que no lo fuera.

Un método más sencillo para controlar el sesgo de la selección es usar a clientes “pivote”, como grupo de control, tal y como recomiendan las herramientas SEEP/AIMS. Debido a que los clientes “pivote” se han seleccionado a sí mismos en el programa, se supone que comparten ciertas características no observables de los clientes existentes. Los clientes pivote han sido utilizados también como grupos de control en diversas publicaciones académicas de estudios de evaluación de impacto. (Buckley 1996a, 1996b; Montgomery *et al.* 1996; Mosely 1996a, 1996b, 1996c, 2001; Mosely and Hulme 1998; and Copestake *et al.* 2001).

El uso de clientes pivote, sin embargo, tiene importantes debilidades metodológicas. Primero, es más apropiado para estudios transversales. Un estudio longitudinal requiere que el grupo de control, o al menos parte de él, no reciba el tratamiento durante todo el periodo de estudio. Es probable, aunque no operativamente viable en la mayoría de los casos, retener los préstamos durante el tiempo de duración del estudio longitudinal de los nuevos clientes que se reclutan a través de las operaciones del día a día. La excepción son los clientes pivote recluidos específicamente como parte de un estudio controlado de evaluación de impacto longitudinal.

En segundo lugar, los clientes pivote no explican por qué algunas personas ingresan antes y otras después. Podría decirse que una persona que entra al programa desde su inicio o al poco tiempo es diferente a alguien que lo hace dos años más tarde. Esto representa un problema menor para los clientes pivote que no tuvieron la opción de ingresar antes, que no lo hicieron antes por razones ajenas a los determinantes de éxito inobservables, o que fueron reclutados en comunidades en las que no hay un programa operando actualmente.⁹ El hecho es que en la mayoría de los casos no sabemos por qué los que entran más tarde han esperado para hacerlo, y casi sin lugar a dudas la explicación suele tener que ver con uno o más rasgos no observables. En resumen, los clientes pivote son un grupo de control que está muy lejos de ser el ideal; la mayoría reconoce que es preferible tener un grupo de personas que no son clientes. Sin embargo, los que apoyan esta propuesta sostienen que es una solución práctica frente a situaciones en las que existen restricciones considerables u obligatorias.

Una propuesta mínima para explicar el sesgo de selección, requiere que los participantes de un grupo de control sean reclutados de la población de microempresarios que son candidatos para ingresar al programa. Es aún mejor si los participantes del grupo de control son reclutados de la población de microempresarios que cumplen con ciertos requisitos de elegibilidad. Desde luego, la debilidad de estas propuestas yace en que suponen, de manera incorrecta, que todos los microempresarios o aquellos que cumplan los requisitos para entrar al programa, lo harían a través de la autoselección. Aun así, es más probable que entren al programa por autoselección los microempresarios que cumplen los requisitos que los que no los cumplen. De esta manera, aunque esta propuesta no elimina el sesgo de selección, al menos tendrá más tendencia a reducir su frecuencia.¹⁰

Hay una gran reserva de microempresarios que son candidatos legítimos o que cumplen con los requisitos de elegibilidad para afiliarse a programas de microfinanciamiento debido al gran número de trabajadores autónomos de escasos recursos que se ganan la vida en el sector informal en LDCs. Pero hay pocas razones justificables para que los investigadores de EI en LDCs no elijan el grupo de control de esta reserva. Por otra parte, los países de la OCDE tienen una incidencia mucho más baja de autoempleo o actividad informal del sector entre las personas de bajos recursos, lo que hace más difícil la selección de grupos de control de entre la reserva de trabajadores autónomos de bajos recursos. De hecho, ninguno de los análisis microempresariales examinados aquí obtiene su grupo de control de esta reserva. Por el contrario, los participantes del grupo de control fueron sacados de grupos de personas que recibieron Ayuda Temporal para Familias Necesitadas (TANF, por sus siglas en inglés) (Raheim and Salome 1999), asistencia social (Raheim and Friedman 1999), prestaciones por desempleo (Kosanovich and Fleck 2002), cupones para alimentos (Departamento de Salud y Servicios Humanos de Estados Unidos, 1994b), y Ayuda para Niños a Cargo (ADFC, por sus siglas en inglés), (Instituto de Desarrollo Social y Económico 1994, Departamento de Salud y Servicios Humanos 1994c). La comparación de los participantes en el programa de microempresas con un ejemplo aleatorio de personas que recibieron ayuda del gobierno, casi con seguridad produce un sesgo de selección y una exageración en el impacto del programa.

La incapacidad de explicar la deserción de los programas

De las evaluaciones de impacto reseñadas, sólo Karlan and Alexander (2002), Himes and Servon (1998) y Mt. Auburn Associates (1998) incluyeron a los desertores del programa entre los que respondieron a la encuesta (sin embargo, ni Himes and Servon ni Mt. Auburn Associates, usaron un grupo de control de personas que no fueran clientes). El hecho de omitir a los desertores del programa del grupo de tratamiento presenta dos fuentes posiblemente serias de sesgo de

selección, a las cuales Karlan (2001) llama *sesgo por muestra insuficiente* y *sesgo por desgaste*. El sesgo por muestra insuficiente surge porque en teoría los desertores tuvieron otra suerte, muy probablemente peor, que la de los que se quedaron. En contraste, el grupo de control (ya sea compuesto de no clientes o clientes pivote) incluye a algunos que triunfan y a otros que fracasan. De ese modo, los estudios de EI que omiten a los desertores del grupo de tratamiento comparan los éxitos del programa con un grupo de control que ha tenido triunfos y fracasos. Por consiguiente, el sesgo por muestra insuficiente produce una exageración sistemática de los impactos del programa.

El sesgo por desgaste aparece si los desertores son sistemáticamente diferentes de los que se quedan, independientemente del impacto. Si, por ejemplo, los participantes ricos tendieran a desertar más que los pobres, entonces el grupo de tratamiento incluiría un porcentaje más alto de participantes más pobres que el grupo de control, y viceversa si la tendencia a la deserción fuera mayor por parte de los participantes más pobres que de los ricos. El resultado es una atenuación sistemática del impacto del programa en el primer caso y una exageración sistemática del impacto del programa en el segundo.

Para probar el efecto que tiene el sesgo por muestra insuficiente y por desgaste en los descubrimientos de evaluación de impacto, Karlan y Alexander (2002), compararon un grupo de tratamiento de una MFI peruana sin los desertores, con un grupo de control de los clientes pivote de la MFI (según las recomendaciones de SEEP/AIMS) y descubrieron pruebas estadísticamente importantes de que hubo impactos positivos. Después de volver a sumar a los desertores al grupo de tratamiento y volver a comparar los grupos, se descubrió que desaparecieron los impactos positivos.

La colocación no aleatoria en el programa

La colocación en el programa no es aleatoria. Es de suponer que los administradores del programa basan la colocación de los participantes en una variedad de criterios estratégicos, como lo son la coherencia con la misión institucional (p. ej., la alta densidad de la población de muy bajos recursos), viabilidad logística (p. ej., a una distancia razonable de la sede del programa), el atractivo económico (p. ej., la alta densidad de la población autoempleada), o la probabilidad de una implementación exitosa (p. ej., una infraestructura relativamente bien creada o un mejor acceso a los mercados). Además, es razonable suponer que las MFI empezarán a expandirse primero hacia aquellos lugares que cumplan mejor con sus criterios estratégicos. Es posible que esos lugares compartan un conjunto de características observables e inobservables que no existirían al mismo nivel en una muestra aleatoria de otros lugares. A su vez, estas características posiblemente sean importantes para explicar los

resultados relativos de la participación en el programa. De esta manera, el control del sesgo de selección también exige que la selección de un grupo de control sea hecha en comunidades que comparten las mismas características observables e inobservables que la comunidad del grupo de tratamiento.

La mejor forma de controlar el sesgo por colocación no aleatoria es haciendo aleatoria la colocación en el programa. Esta propuesta quizá parezca poco práctica a nivel administrativo, pero no tiene por qué serlo, especialmente si está diseñada para aprovechar los límites naturales de la expansión del programa. Digamos, por ejemplo, que una MFI planea expandirse a una cantidad determinada de lugares en los siguientes dos años. Todos esos lugares cumplen con los criterios de colocación del programa. No obstante, debido a las restricciones de los recursos, el programa no se puede expandir a todos los lugares que quiere de una sola vez, así que planea pedir la expansión a una parte de los lugares este año, y el próximo a todas. Ya que la MFI en gran medida es indiferente al orden de la expansión, si se establece un orden aleatorio en el proceso de selección, se cumple con los criterios de expansión del programa y los controles del sesgo por colocación no aleatoria.

Los únicos ejemplos de colocación aleatoria en programas entre los estudios de EI reseñados aquí, son MkNelly y Dunford (1999a, 1999b) en evaluaciones de programas de microfinanciación en Bolivia y Ghana. Ambas evaluaciones siguieron la misma metodología. La administración del programa eligió comunidades a las cuales expandirse en los siguientes dos años. Después, el personal del programa visitó cada comunidad para reclutar participantes para el programa. En todos los casos, el personal del programa dejó claro que la comunidad podría ser asignada al grupo de control que no recibiría el programa sino hasta dos años después. Luego, el personal del programa recogió información de base de todas las personas elegibles que optaron por ingresar al programa. Después de reunir la información, los investigadores estratificaron las comunidades de estudio según sus características comunitarias clave, tales como su tamaño, acceso a la vía principal, distancia de un mercado, y acceso al agua. Por último, los investigadores asignaron aleatoriamente comunidades que controlar y comunidades de tratamiento, de forma que se minimizaran las diferencias entre las características comunitarias clave.

Coleman (1999, 2001b) usó un enfoque un tanto diferente en su evaluación de dos programas de microfinanciamiento en Tailandia. La comunidad de tratamiento incluyó a tres pueblos con acceso a los programas con los que trabajó a lo largo de dos a cuatro años. Se preseleccionaron seis comunidades de control para recibir el programa un año después de que se les identificara. Los habitantes de las comunidades de control eligieron por sí mismos si participarían en el programa. Para justificar la posibilidad de que el orden en el que los catorce pueblos recibirían el programa no fuera aleatorio, Coleman reunió una tercera

muestra de no participantes en cada uno de los pueblos. Coleman descubrió que los cálculos ingenuos que no explicaban la autoselección o colocación no aleatoria en el programa, sobreestimaban considerablemente el impacto del programa en varias variables de resultados, incluyendo la riqueza general, propiedades prediales, activos agrícolas que no son propiedades, ahorros, e ingreso familiar. Llegó a la conclusión de que los factores condicionantes del ingreso de negocios pequeños eran las características inobservables, no la falta de participación en programas de microfinanciamiento.

También se pueden usar técnicas económicas para controlar el sesgo de selección. Un ejemplo es el enfoque usado originalmente por Pitt and Khandker (1994) para evaluar el impacto de tres programas en Bangladesh (y posteriormente usado por McKernan 1996; Pitt et al. 1998; Pitt and Khandker 1998; Khandker et al. 1998; Pitt et al. 1999; Morduch 1998; y Khandker 2001). Su enfoque aprovechó las reglas del programa que excluían a las familias cuyos activos fijos obtenidos a través de la participación en el programa superaban cierta cantidad. De hecho, compararon los resultados entre familias elegibles y no elegibles en pueblos en los que existía el programa, con los resultados entre familias elegibles y no elegibles en pueblos en los que no había programas, y luego compararon las diferencias entre ellas. Atribuyeron cualquier discrepancia entre estas dos diferencias a la participación en el programa. Al igual que Coleman, su estudio demostró que los cálculos ingenuos que no logran explicar el sesgo de selección, sobreestiman considerablemente el impacto (entre otros estudios de EI que usan técnicas econométricas para controlar el sesgo de selección están Lapar, Graham, and Meyer 1995; Lapar et al. 1995; Zeller et al. 1996; y Zaman 2001). El inconveniente de los enfoques econométricos como el de Pitt y Khandker es la complejidad de sus métodos econométricos, que sólo están accesibles para un grupo relativamente pequeño de metodólogos igualmente sofisticados, pero no para la mayoría de las autoridades y casi con certeza tampoco para el experto promedio.¹¹

Sesgo por contaminación

El sesgo por contaminación ocurre cuando el grupo de control se contamina por el contacto con el grupo de tratamiento. La contaminación puede ocurrir de varias formas; por ejemplo, si los miembros del grupo de control conocen a los miembros del grupo de tratamiento, algunos miembros de ambos grupos tienen conocidos en común, el programa inicia contacto con miembros del grupo de control o viceversa, o la información sobre el programa se difunde a través de redes sociales formales o informales. La contaminación también ocurre cuando la participación en el programa crea factores externos positivos o negativos que influyen en el comportamiento o resultados de los no participantes.

Una vez que los participantes del grupo de control son contaminados, el

investigador no puede volver a tener la certeza de si han sido influenciados o cómo, a consecuencia el comportamiento del grupo de control y otros resultados. Según el resultado, ya sea mejor o peor, que produzca la contaminación entre los participantes del grupo, se creará una atenuación o exageración de los impactos reales del programa. El sesgo por contaminación puede ser manejado de una forma suficientemente fácil, ubicando al grupo de control lejos del grupo de tratamiento. Entre más lejos estén, menos probabilidades habrá de contaminación (aunque esto debe ser sopesado contra el aumento en costo y problemas logísticos).

Sesgo de recuerdo

La investigación de encuestas exige que los entrevistados recuerden información sobre sus condiciones socioeconómicas, comportamientos, actitudes y relaciones sociales. Hay varias razones por las que puede haber sesgo en las respuestas recordadas. Los encuestados muestran tendencias naturales y razonables de 1) querer agradecer al entrevistador, hablar bien de sí mismos, o evitar revelar información vergonzosa, 2) sospechar de los motivos del encuestador, 3) jugar con el proceso, 4) hacen conjeturas al azar, o 5) dar cualquier respuesta para evitar que se prolonguen las invasiones no deseadas a su tiempo. En general, entre más personales o invasivas son las preguntas (p. ej., prácticas sexuales), o las más difíciles de calcular (p. ej., ingresos familiares), más pueden esperar los investigadores que se cuelen las imprecisiones y el sesgo en las respuestas.

Si las entrevistas se hacen a través del personal del programa, tal y como lo recomiendan las herramientas de SEEP/AIMS, se aumenta considerablemente el riesgo de sesgo en las respuestas, y usar al personal para entrevistar a sus propios clientes prácticamente lo garantiza. Esto no pretende indicar que no deba usarse el personal del programa. De hecho, la utilización del personal del programa para realizar las investigaciones a menudo es una concesión práctica y necesaria frente a las restricciones del programa. Además, integrar EI o investigaciones del mercado a los sistemas del programa o deshabituarse de traer evaluadores externos, puede bien exigir el uso de personal de campo para recaudar información de impacto. Sin embargo, sólo debe hacerse uso del personal de impacto para entrevistar a sus propios clientes como último recurso. Los programas que usan personal de campo para efectuar investigaciones de impacto deben estar totalmente informados de los riesgos que esto supone y preparados para hacer lo posible para mitigar esos riesgos. El mejor enfoque que se puede tomar para disminuir estos riesgos es combinar fuertes dosis de entrenamiento, monitoreo y verificación de información.

Incluso en los casos en que los encuestados se sienten inclinados a dar respuestas de buena fe, es posible que recuerden o no recuerden la información con una exactitud razonable. Entre más tiempo pase, más difícil es calcularlo con

precisión. La estacionalidad también puede causar estragos, ya que las respuestas varían dependiendo de la época del año. Las estrategias para mitigar esta fuente de sesgo son usar periodos de tiempo más cortos y encuestar en diferentes épocas del año para reflejar las temporadas clave o ciclos de cultivo (ver, por ejemplo, Coleman 1999, 2001b; Pit et al., 1998; y Mustafa et al., 1995).

La fungibilidad en los préstamos

Los préstamos que reciben los participantes del programa suelen entremezclarse con otras fuentes de ingresos familiares, para ser gastados según las necesidades de sustento y prioridades de gastos de la familia. En otras palabras, los préstamos no necesariamente están destinados a ser invertidos en las empresas de los participantes. De forma similar, los fondos para el pago de los préstamos no necesariamente salen del flujo de efectivo de la empresa, sino muy a menudo de otras fuentes de ingresos familiares. La fungibilidad en los préstamos complica enormemente la capacidad de EI de formar una conexión directa entre la cantidad del préstamo y los cambios en el ingreso, consumo, acumulación de activos, facultamiento individual, etcétera. El estado actual de la metodología del análisis de impacto no incluye procedimientos establecidos para explicar la fungibilidad de los préstamos.

No obstante, la fungibilidad se vuelve un problema menor si el enfoque de EI es la cartera económica de la familia (Chen and Dunn 1996). El concepto de la cartera económica de la familia reconoce explícitamente la fungibilidad del dinero como un componente vital del sustento y estrategias de reacción de las familias. Le interesa menos la ruta de la cadena causal de impacto, *per se*, que la forma en que terminan manifestándose los impactos en cada nivel de análisis.¹²

Otro problema causado por la fungibilidad de préstamos es que incluso las respuestas sinceras a las preguntas de la encuesta son engañosas. Para ilustrar esta idea, Coleman (2001a) pone el ejemplo de una participante de un programa que usa sus préstamos del programa para pagar los costos de la escuela de sus hijos en lugar de vender sus activos para cubrir esos gastos, como suele hacerlo. En este caso, el beneficio de incremento real de la participación en el programa para esta mujer es conservar sus activos, lo que no está medido, y no el pago de las matrículas, cosa que está medida.

De los estudios reseñados aquí, sólo Diagne (1998) y Zaman (2001), hablan explícitamente de la fungibilidad de préstamos. La propuesta de Diagne fue burlar el problema haciendo que el tratamiento relevante fuese el acceso al crédito en lugar de la entrega del crédito, basándose en el razonamiento de que los cambios en los resultados debido al acceso al crédito eran más fáciles de aislar e identificar que los cambios en los resultados debido a la entrega del crédito. Zaman, por otra parte, usó un enfoque econométrico basado en el modelo de cartera económica.

Su modelo da por supuesto que el dinero prestado se gasta según las necesidades de hogares que maximizan sus servicios públicos y que, si se controlan otros factores a través de la aplicación de los procedimientos econométricos, es posible atribuir resultados específicos a la entrega del préstamo. Otro enfoque posible para explicar la fungibilidad del préstamo es reunir información sobre el uso de los fondos y fuentes de pago del préstamo a través de preguntas adicionales en la encuesta o investigación cualitativa.

La evaluación longitudinal contra la transversal

Es de suponer que los impactos ocurren con el tiempo, y que las evaluaciones longitudinales que dan seguimiento a los resultados a lo largo de diferentes momentos arrojan más luz sobre este proceso y que las evaluaciones transversales que se llevan a cabo en un sólo momento. Si no es posible realizar un estudio longitudinal, los investigadores pueden autorizar un periodo de tiempo en el que se realice una evaluación transversal, eligiendo intencionalmente a participantes con ciertos años específicos de experiencia en el programa, y comparándolos ya sea con un grupo de control de no clientes o de clientes pivote. Por ejemplo, las herramientas SEEP/AIMS recomendaron que los investigadores elijan participantes con uno y dos años de experiencia en el programa como participantes del grupo de tratamiento, suponiendo que estos representan momentos útiles en los que se pudieron haber observado impactos considerables.¹³

Las virtudes de las evaluaciones de impacto transversal son el costo relativamente bajo y las bajas exigencias en cuanto a la recopilación de información. En los estudios longitudinales, el desgaste del panel (cuando los participantes de los grupos de tratamiento y control renuncian al estudio), también puede representar un problema.¹⁴ Las formas más comunes de manejar el desgaste del panel son dar seguimiento a los desertores del programa o hacer un número mayor de encuestas de lo que sería necesario de otra forma, situaciones que añaden más costos relativos a los estudios transversales.

A pesar de que existe una preferencia por las evaluaciones longitudinales, varios estudios de impacto transversal han sido publicados por investigadores muy respetados o en revistas académicas muy prestigiadas, y las evaluaciones transversales integran el pleno de las evaluaciones de programas de microfinanciamiento; 63 evaluaciones de microfinanciamiento fueron transversales, a comparación de sólo 27 longitudinales. Por otra parte, 11 de las evaluaciones de microempresas reseñadas fueron longitudinales y 9 transversales. Cuando no son factibles las evaluaciones longitudinales, los estudios transversales de cada caso son una segunda alternativa aceptable. No obstante, los mejores procedimientos sostienen que, deben llevarse a cabo evaluaciones longitudinales siempre que sea factible.

Otras deficiencias metodológicas

La sección anterior reseñó las dificultades metodológicas más comunes en las evaluaciones basadas en el impacto. Es esencial comprender estas dificultades, sus causas, y sus soluciones para que las autoridades evalúen la validez de los estudios de EI para así poder determinar su utilidad a la hora de informar la toma de decisiones de orden público. Estas deficiencias metodológicas tienen menos que ver con temas de precisión científica que con dar a las autoridades un análisis profundo de los beneficios y costos de la participación en el programa.

Lo ideal es que los fondos públicos sean destinados a aquellos programas sociales que ofrezcan el bienestar social neto más alto, definiendo el bienestar social neto como el valor actual de los beneficios del programa menos el valor presente de los costos del programa. A su vez, calcular el bienestar social supone lo siguiente: 1) la identificación de todos los beneficios y costos relevantes del programa, 2) el cálculo de los beneficios y costos relevantes del programa, 3) la conversión de beneficios y costos relevantes a unidades estandarizadas para permitir la comparación dentro y a través de los programas y 4) el cálculo de los beneficios y costos relevantes del programa para reflejar valores y prioridades sociales. Por contraste, la mayoría de los estudios de EI dan sólo una imagen muy limitada de los beneficios del programa y prácticamente ninguna información sobre el costo del programa; no tratan de convertir los beneficios en unidades estandarizadas, y consideran como iguales a todos los beneficios y costos. Cada una de estas deficiencias y sus repercusiones serán abordadas brevemente más adelante.

La omisión de los beneficios del programa

Todas las evaluaciones de impacto toman decisiones sobre cuáles beneficios examinar. La mayoría examinan alguna combinación de beneficios a nivel de empresas y familiares, muchos también examinan los beneficios a nivel individual, pero casi todos pasan por alto los beneficios a nivel comunitario, aunque las pruebas indican que pueden ser importantes en conjunto.¹⁵ La elección de los beneficios examinados puede estar basada en una serie de criterios, por ejemplo, prioridades del donante o programa, convención de la industria, o interés personal. Pero cualquiera que sea el criterio, el resultado final es que los beneficios importantes inevitablemente son omitidos de las evaluaciones del programa.

La omisión de beneficios importantes en un programa da a las autoridades una imagen incompleta del impacto del mismo. Más allá de esto, la única forma en que las autoridades pueden formarse una imagen completa del impacto del programa es improvisando los descubrimientos de evaluaciones de varios programas que examinan diferentes tipos de beneficios a diferentes niveles de

análisis. A pesar de que este enfoque puede ser útil, pues lo mejor que puede hacer es dar una imagen muy amplia del impacto, su utilidad para evaluar el impacto de un programa en especial es limitada, debido a la considerable disparidad contextual bajo la cual operan los programas.

Omisión de costos del programa

La mayoría de las evaluaciones de impacto no mencionan, ni mucho menos calculan los costos del programa. Entre los costos relevantes del programa están el valor actual de los costos administrativos y el costo monetario y de oportunidad de los fondos donados o invertidos, pasivos a largo plazo, ayuda en especie, costos de operación basados o no en el precio que son producidos por los participantes del programa, y costos de sustitución (los beneficios contraídos por los participantes del programa a expensas de los que no participan). Calcular los costos directos del programa y los costos del precio para los participantes es relativamente sencillo, pero calcular la ayuda en especie, costos de operación que no están basados en el precio y costos de sustitución, será una carga hasta para el investigador más concienzudo, lo que quizá explica por qué tan pocos lo han hecho.

Como ejemplo están los costos de operación no basados en el precio. Los programas de préstamos a grupos por lo general exigen que los participantes formen los grupos, participen en reuniones de grupo dos o cuatro veces al mes, y monitorean y hacen cumplir el desempeño del préstamo al grupo. Tales costos de transacción no basados en el precio son difíciles de medir, aunque imponen cargas considerables para los participantes del programa. Los costos de sustitución son aún más difíciles de medir, y pueden ocurrir, por ejemplo, cuando los programas atraen a un gran número de personas autoempleadas a los sectores que son atractivos para los microempresarios no especializados y caracterizados por tener pocos obstáculos para ingresar, mucha competencia, y pocas ganancias. Los costos de sustitución pueden ser considerables, y en algunos casos alcanzan a ser la mitad de los beneficios netos que están acumulando los participantes del programa (Bendick and Egan, 1987). De las evaluaciones de impacto reseñadas, sólo Kosanovitch y Fleck (2002) tienen en cuenta los costos de sustitución.

Los costos del programa posiblemente deberían ser incluidos en las evaluaciones de impacto hasta el grado en que estén disponibles o puedan ser calculados de forma razonable, al menos para ofrecer una base de comparación, como el costo por resultado. Si los costos no se pueden calcular razonablemente, se puede esperar al menos que los investigadores identifiquen los posibles costos, quizá den algún cálculo de su orden de magnitud, y expliquen cómo podrían afectar el análisis en el que serán incluidos.

Kilby and D'zmura (1985) es la única evaluación de un programa de

microfinanciamiento que en la reseña demostró haber conducido un análisis de beneficio y costo. Kilby y D'Zmura evalúan principalmente los beneficios directos e indirectos según son medidos por el valor agregado a las empresas asistidas y el valor agregado a las empresas fuera del proyecto. El primero incluye sueldos, alquiler, interés, y ganancias modificadas según el costo de oportunidad de trabajo,¹⁶ y el segundo de compras de factores de producción realizadas por empresas asistidas y compras de bienes de consumo efectuadas con ingresos de factor directo obtenidos por empresas que recibieron ayuda. Entre los otros beneficios tomados en cuenta están la capacitación, reducción de precio, beneficios de desvío (los beneficios derivados del desvío de un préstamo de microempresas a otro fin), y los sueldos calculados para los muy pobres. Entre los costos valorados están todos los gastos administrativos, deudas incobrables, y merma de capital (la tasa de interés efectiva por debajo de la tasa de inflación).

Varias evaluaciones de programas de las microempresas reseñadas realizaron al menos algún tipo de análisis de beneficio y costo. Else and Clay-Thompson (1998) calcularon el costo por unidad de producción para los clientes atendidos, empleos y negocios creados, negocios que recibieron ayuda, y préstamos hechos. De forma similar, el Departamento de Salud y Servicios Humanos de Estados Unidos (1994), comparó el aumento en ganancias por cupones de alimentos de los participantes de algunas pruebas, con los costos administrativos del programa y los aumentos en pagos de ayuda pública a los participantes de las pruebas.

El análisis más completo de beneficios y costos llevado a cabo entre las evaluaciones reseñadas de microempresas, es la evaluación de Kosanovich and Fleck (2002) de Programas de Ayuda para el Autoempleo en Maine, Nueva Jersey y en Nueva York. Los autores evaluaron los costos y beneficios para los participantes del programa, gobiernos del estado, y no participantes. Entre los beneficios evaluados para los participantes estuvo la ganancia de ingresos gracias al autoempleo o empleo por salario o sueldo, desarrollo profesional, satisfacción laboral, y desarrollo económico comunitario. Entre los beneficios del gobierno que fueron evaluados están el aumento en los ingresos por los impuestos y la reducción en transferencias de asistencia social. Los costos para los participantes evaluados fueron costos económicos de administración, capacitación, asesoría del programa y los costos de oportunidad para los participantes que renuncian a la búsqueda de empleo y a la posibilidad de volver a ser contratados mientras están autoempleados. Los costos del gobierno evaluados fueron los pagos de seguro, administración del programa, capacitación y asesoría. Por último, entre los impactos de sustitución que fueron evaluados estuvieron los cambios en los beneficios de seguro por desempleo y pagos de impuestos que benefician a los participantes pero que son pagados por no participantes, y la reducción en pagos de asistencia social o seguro por desempleo a largo plazo que beneficiaron a los no participantes pero que fueron pagados por los participantes.

La falta de estándares comunes para realizar una comparación

La comparación de beneficios y costos dentro y a través de varios programas (ya sea con otro programa de microfinanciamiento o microempresas, o con algún otro tipo de programa de desarrollo o empleo) exige que los beneficios y costos sean convertidos a unidades estandarizadas. Por lo general, esto implica la conversión a unidades monetarias. Sin algún estándar común, las autoridades no tienen otra forma además de sus propios cálculos aproximados para comparar, por ejemplo, el valor de un empleo creado con un aumento en la autoestima del participante en un programa, o el valor de los trabajos creados por el mismo programa con el valor del aumento en el acceso del participante al cuidado médico en otro programa. La estandarización de beneficios y costos tiene la ventaja adicional de obligar a los investigadores a ser explícitos sobre sus suposiciones y juicios, de ese modo mejoran el análisis y facilitan las evaluaciones *ex post* del rigor metodológico.

La incapacidad de sopesar los beneficios y costos

Las evaluaciones de impacto no distinguen entre el valor relativo de los resultados, sino que tratan a todos los beneficios y los costos del programa como iguales. Los beneficios tan dispares como el aumento en el consumo familiar y el aumento en la autoestima de los participantes, por ejemplo, son tratados como iguales, así como los beneficios similares, pero claramente distinguibles, como los empleos de tiempo completo y medio tiempo que han sido creados. De hecho no todos los beneficios ni todos los costos son iguales. La sociedad otorga un valor más alto a algunos beneficios o costos del programa que a otros. Por lo tanto, posiblemente las evaluaciones de impacto deban reflejar valores sociales a través de algún tipo de sistema de peso. Algo que no es un sistema de peso explícito, pero sería de gran ayuda, sería algún diálogo sobre el peso relativo y cómo pueden afectar el análisis. No obstante, en la evaluación de impacto reseñada no se encuentran ni los sistemas de peso ni tipo de diálogos sobre el peso relativo de los resultados de un programa.

Otras deficiencias metodológicas varias

Las evaluaciones de impacto están sujetas a varias otras deficiencias metodológicas. Estas deficiencias metodológicas están relacionadas en su mayor parte con las decisiones tomadas por investigadores de EI sobre qué información reportar y cómo hacerlo. Las decisiones que han sido tomadas acerca de la presentación de información, pueden influir materialmente en los resultados de la evaluación y en cómo son interpretados esos resultados.

Las decisiones tomadas acerca la presentación de la información reflejan una serie de factores, incluyendo la destreza del investigador, los objetivos de la evaluación, las decisiones subjetivas sobre la relevancia de la información, o los sesgos del investigador o el que proporciona los fondos. Aunque no podemos suponer que las decisiones sobre la presentación de la información han recibido necesariamente la influencia del sesgo del investigador o proveedor de fondos, ni podemos asumir que nunca hubo sesgo en las decisiones. Este tema toma más relevancia cuando nos damos cuenta que en muchos casos las evaluaciones de impacto reciben fondos o son conducidas por gente u organizaciones que apoyan abiertamente el desarrollo microempresarial. Esto en combinación con todos los otros factores que influyen en las decisiones de presentación de información, sugieren que es útil que las autoridades estén familiarizadas con las formas en las que tales opciones pueden influir en los resultados de evaluación. Schreiner (2002) cataloga varios de los ejemplos más comunes encontrados en las evaluaciones de programas de microempresas en Estados Unidos. Sin embargo, las mismas prácticas se pueden encontrar a mayor o menor grado en otras evaluaciones de programas de microfinanciamiento o microempresas, de tal forma que es útil resumir los argumentos de Schreiner a continuación.

Presentación selectiva de resultados del programa. A veces las evaluaciones de impacto presentan selectivamente información de forma que sesga el análisis y las conclusiones. Un ejemplo de esto es no poder distinguir entre valores (p. ej., préstamos pendientes o aprendices actuales) y flujos (p. ej., préstamos desembolsados o gente capacitada). Los flujos reúnen el desempeño anterior con el actual, de ese modo producen una imagen tergiversada del desempeño. Los flujos también exceden a los valores en cualquier momento, de tal forma que reportar el flujo en lugar del valor nos da una visión más favorable del desempeño del programa. Cuál de los dos es apropiado reportar depende de la pregunta que se haga. A pesar de eso, los investigadores deben reportar la unidad que está siendo utilizada y la razón por la que se está utilizando.

Una práctica similar es reportar las cifras en conjunto en lugar de las cifras en conjunto modificadas según el número de participantes del programa. Por ejemplo, determinar que los participantes de un programa lograron ventas de 3.5 millones de dólares, suena más impresionante que determinar que las

ventas promedio de cada participante del programa dieron un total de doce mil dólares.

Otro caso de presentación selectiva es el uso de estadísticas. La interpretación de descubrimientos puede llevar la influencia del “giro” que se les dé. Por ejemplo, reportar que “más de la mitad” de los participantes del programa aumentaron los rendimientos de su empresa es darle un giro positivo al descubrimiento más negativo de que “casi la mitad” de los participantes del programa no logró aumentar los rendimientos de su empresa. Reportar las estadísticas del resumen general cuando hay estadísticas desagregadas disponibles, es otra práctica encontrada en las evaluaciones de impacto. Por ejemplo, el índice de pago de préstamos es una medida que se reporta comúnmente para indicar la calidad de la cartera; no obstante, esta es una estadística del resumen general que oculta información crucial visible en medidas mucho mejores, como una cartera en riesgo vencida por antigüedad.¹⁷

Un último ejemplo de la presentación selectiva es no reportar las deserciones del programa. Las deserciones quizá sean la forma más sencilla y eficaz de determinar si el programa crea un valor para los participantes. Es razonable llegar a la conclusión de que los programas con altos índices de movimiento de clientes son menos eficaces para crear un valor y tienen menos impacto que los programas con índices relativamente bajos de movimiento de clientes. Además, la exclusión de deserciones del análisis de impacto casi con seguridad produce una exageración sistemática de los beneficios del programa.

Cálculos equivocados de los beneficios del programa. Es común que las evaluaciones de impacto hagan un cálculo equivocado de los beneficios del programa. Schreiner señala, por ejemplo, que algunas evaluaciones de microempresas reportan el inicio de nuevos negocios como si todos fuesen atribuibles directamente a la participación en el programa, pasando por alto que algunos participantes ingresan al programa con empresas ya existentes, otros empiezan negocios una vez que han desertado, y la mayoría de las nuevas empresas fracasan en pocos años.

Otro ejemplo es hacer un reporte equivocado del ingreso de la empresa, como la práctica de 1) reportar niveles de ingresos en lugar de cambios de ingreso, 2) reportar ingresos de la empresa en lugar de rendimientos de la empresa, y 3) no determinar los ingresos claramente. En el primer caso, reportar niveles de ingreso exagera los beneficios del programa a no ser que el ingreso fuera nulo antes de integrarse al programa. En el segundo caso, el ingreso de la empresa da una imagen tergiversada del impacto porque no explica el rendimiento según el tiempo trabajado y el capital invertido. Los niveles de ingreso absolutos que parecen altos quizá no lo parezcan una vez modificados según el tiempo trabajado y el capital invertido. Por último, la falta de claridad sobre la definición de ingresos

permite que los investigadores reporten una variedad de resultados como el ingreso de la empresa, algunos más favorables que otros. Un ejemplo perfecto es la práctica de reportar ingresos transitorios de la empresa como ingresos o como un valor sustitutivo de ingresos, lo que inevitablemente distorsiona los descubrimientos superiores.

Cómo juzgar la metodología de EI

El conocimiento de los obstáculos y deficiencias metodológicas de la evaluación de impacto es útil para juzgar la calidad de las evaluaciones del programa. Pero antes de hacer un juicio, debe repetirse que aunque la pureza metodológica es un ideal loable, pocas veces se logra en la práctica. Casi todas las evaluaciones de impacto sufren de alguna deficiencia metodológica u otra. Algunas son evidentes, y otras requieren una lectura más meticulosa para ser vistas. Algunas son resultado de restricciones ambientales o de recursos, otras de decisiones subjetivas tomadas por los investigadores, y otras de ambas cosas.

En la práctica, la mayoría de los investigadores se ven obligados a hacer concesiones de algún tipo ante las restricciones de recursos o ambientales, y deben tomar decisiones subjetivas sobre cómo manejarlas. Desde un punto de vista más extenso, todos los investigadores deben tomar un gran número de decisiones subjetivas sobre cómo diseñar e implementar estudios de evaluación, analizar información de evaluación y reportar descubrimientos de evaluación.

Aunque hay pocas discrepancias entre los investigadores sobre qué constituye una cuestión o un obstáculo metodológico importante, o qué constituye una metodología ideal, hay diferencias considerables sobre qué es una metodología aceptable, cuáles son los sacrificios aceptables entre el rigor y el costo, qué concesiones metodológicas se pueden hacer de forma legítima a las restricciones ambientales o de recursos, o cuáles decisiones subjetivas son válidas y cuáles no. La publicación en revistas académicas reseñada por sus pares, de ninguna manera supone la aprobación metodológica. Los consejos editoriales de revistas académicas tienen estándares metodológicos diferentes, al igual que los críticos que utilizan. Pocas evaluaciones académicas de impacto publicadas y reseñadas por colegas cumplen con los estándares de todos en cuanto al rigor metodológico.¹⁸ Pero si el golfo metodológico entre los investigadores académicos es amplio, el golfo entre los académicos y expertos es un verdadero abismo. No es decir que a los expertos no les importe la validez científica, hasta cierto grado sí les importa. Pero es un lujo que la mayoría de ellos siente que no puede permitirse. Se conforman con dejar que los universitarios determinen los detalles metodológicos, mientras se concentran en dirigir programas. Además, ¿qué administrador de una organización ha tomado decisiones clave en el programa con un grado de certeza de +-.05?

Teniendo en cuenta las discrepancias sobre el rigor metodológico adecuado, no debe esperarse que las autoridades solucionen lo que no pueden solucionar los investigadores profesionales y expertos. Sin embargo, lo que sí pueden hacer es insistir en que se haga una divulgación completa de los enfoques y deficiencias metodológicos, las restricciones que se enfrentan y los sacrificios y las decisiones subjetivas que se toman. La divulgación completa fomenta la transparencia y la transparencia fomenta los juicios bien fundamentados sobre la idoneidad metodológica y la interpretación bien fundamentada de los descubrimientos de las evaluaciones. La divulgación completa es un objetivo fácil, pero crítico, al que deben aspirar todas las evaluaciones de impacto. Como señaló Schreiner (2002): “El centro del método científico social no son los experimentos sino lo explícito”. (p. 69).

Recomendaciones

Si se toma todo lo anterior en cuenta, se proponen las siguientes recomendaciones para ayudar a que las autoridades juzguen la calidad o rigor de las evaluaciones de impacto e interpreten sus descubrimientos para así orientar sus esfuerzos de mejorar las prácticas de EI y su utilidad.

Aliente el uso de enfoques de método combinados. La fuerte dependencia en las encuestas de impacto probablemente refleje tanto o más la capacitación, inclinaciones y sesgos de los investigadores de EI que cualquier superioridad metodológica inherente. Los métodos cualitativos (ya sean basados en la tradición de humanidades o PLA) tienen mucho que ofrecer, especialmente en términos de producir una comprensión más profunda y matizada del impacto. Los métodos cualitativos también pueden lograr mucho para superar los problemas de la fungibilidad de préstamos, y no han sido usados suficientemente con este objetivo. Una de las evaluaciones de impacto más intuitiva que ha sido reseñada fue el estudio etnográfico de Todd (1996) de los miembros del banco Grameen; ofrecía una riqueza de comprensión que no tenían la mayoría de los análisis basados en encuestas. Cuando no hay suficiente tiempo o recursos para permitir estudios etnográficos de este tipo, ciertos métodos de evaluación rápida como PLA, pueden dar de forma similar una información perspicaz y útil sobre el proceso a través del cual ocurre el impacto y cómo se manifiesta a diferentes niveles de análisis. La encuesta de impacto debe seguir siendo la herramienta principal, pero no necesita ser tan dominante como lo ha sido hasta la fecha.

Use grupos de control válidos. Si el objetivo de una evaluación basada en una encuesta es atribuir un impacto a la participación del programa, se debe usar un grupo de control válido. No tiene mucho sentido invertir los escasos fondos públicos en una evaluación de impacto que no pueda esperar responder las preguntas que se le hagan. Como mínimo, un grupo de control válido debe

estar integrado por microempresarios elegibles para ingresar al programa. Los clientes pivote cumplen con este requisito mínimo cuando existen restricciones obligatorias, pero se deben hacer esfuerzos para elegir los clientes pivote que no han tenido la oportunidad de ingresar al programa antes, como lo son los clientes pivote de comunidades a las que se acaba de expandir el programa.

Controle el sesgo de selección. Se debe hacer todo esfuerzo razonable para controlar todas las formas de sesgo de selección. La asignación y la colocación aleatorias en el programa deben ser utilizadas tanto como sea posible. La asignación aleatoria no tiene por qué ser costosa o demasiado intrusiva. Coleman (2001a), por ejemplo, calcula que la reproducción de su metodología podría ser implementada fácilmente a un costo de sólo 25 000 a 50 000 dólares, lo que se acerca al alcance mínimo que de todas formas costaría una evaluación de impacto de alto nivel. El estudio de Benus et al. (1994) demuestra que la asignación aleatoria es posible también en un contexto de la OCDE. Incluso cuando controlar el sesgo de selección cuesta más, vale la pena el incremento adicional de gasto. Una vez más, no tiene mucho sentido invertir dinero en una evaluación de programa para la que se pueda predecir antes de tiempo que dará resultados cuestionables o inválidos, especialmente cuando los descubrimientos válidos serían producidos gracias a una inversión más grande. Los investigadores también deben mostrar pruebas de controlar el sesgo de selección y hablar francamente sus consecuencias.

Incluya a los desertores del programa en el grupo de tratamiento. La exclusión casi universal de desertores del programa de los grupos de tratamiento es una deficiencia metodológica seria que casi con toda seguridad ha producido una exageración sistemática de impacto. Incluir a los desertores del programa requiere una inversión adicional de dinero y esfuerzo, pero ciertamente se puede hacer en la mayoría de los casos, y los beneficios en términos de precisión estadística probablemente hagan que la inversión adicional realmente valga la pena.

Ubique el grupo de control a una distancia suficiente del grupo de tratamiento. Para evitar el sesgo por contaminación, el grupo de control debe ser ubicado a suficiente distancia, dentro de la lógica, del grupo de control, para minimizar la probabilidad de sesgo por contaminación.

Destine suficiente dinero para realizar evaluaciones metodológicas de impacto firmes. Si las autoridades o agencias que proporcionan los fondos exigen que los programas realicen evaluaciones de impacto, se deben destinar suficientes fondos para implementarlas y hacerlo de una forma metodológica firme. Muy a menudo las agencias esperan pruebas convincentes del impacto pero no destinan suficiente dinero (o a veces nada) para la implementación de una evaluación de impacto válida. Lo único que esto tiende a producir es un esfuerzo a medias y cinismo sobre la evaluación de impacto.

Realice evaluaciones longitudinales cuando sea factible. Las evaluaciones transversales son aceptables, pero deben ser una segunda opción clara y deben estar reservadas para situaciones en las que no sean factibles las evaluaciones longitudinales.

Póngase de acuerdo con antelación sobre el alcance claro y completo de trabajo con la administración del programa y los investigadores de EI. Pueden surgir problemas metodológicos porque las expectativas o malentendidos difieren entre las agencias que proporcionan los fondos, la administración del programa y los investigadores. Así, probablemente se pueden evitar muchos problemas metodológicos si las agencias que proveen los fondos colaboran más de cerca con la administración del programa y los investigadores, para definir un alcance de trabajo que cubra, como mínimo, a) los objetivos de la evaluación, b) las opciones metodológicas que sean coherentes con los objetivos de la evaluación, c) los niveles de impacto que deberán ser evaluados y los indicadores correspondientes que serán usados, d) las restricciones de campo (objetivos y la metodología quizá deban ser negociados en vista de las restricciones de campo), y e) los requisitos de informes (qué debe reportarse y cómo).

Realice análisis de beneficio y costo más rigurosos. Aún no se ha determinado si el desarrollo microempresarial merece un gran gasto de dinero público debido a la escasez de buenos análisis de beneficios y costos. Aunque es útil saber si y hasta qué grado los programas de microfinanciamiento y microempresariales benefician a los participantes y a los que no participan, esta información no necesariamente permite respuestas bien fundadas sobre si se debe o hasta qué grado otorgar fondos a los programas con relación a otras opciones de políticas públicas. Esto exige evaluaciones rigurosas del programa que analicen la variedad de beneficios y costos. También exige algún método para estandarizar y sopesar los descubrimientos.

Schreiner (2002) señala que la importancia del análisis riguroso de beneficios y costos crea un posible dilema del prisionero para los programas de microfinanciamiento y microempresa, en el sentido de que un mayor rigor (y divulgación) los coloca en una desventaja entre las autoridades en relación con las políticas de competencia que no se adhieren a los altos niveles correspondientes de rigor en la evaluación. De modo que una recomendación secundaria es que las autoridades encargadas de crear políticas sometan a todas las evaluaciones de programas a estándares altos de rigor metodológico. A pesar de eso, la mala metodología de evaluación en programas sociales en competencia no libera a los programas de microfinanciamiento o microempresas de la responsabilidad ética de hacer esfuerzos razonables y válidos para justificar los escasos fondos públicos que reciben.

Insista en la divulgación completa y sincera. Los investigadores de EI deben ser

responsables de divulgar totalmente y sinceramente toda la información relevante para que los que no son expertos entiendan e interpreten los descubrimientos de la evaluación. Entre las cosas que deben ser divulgadas están, como mínimo, a) las metodologías elegidas, porqué fueron elegidas, y sus consecuencias, b) las debilidades de las metodologías elegidas y sus consecuencias, c) las restricciones enfrentadas y los sacrificios hechos y sus consecuencias, d) las decisiones subjetivas tomadas y sus consecuencias, e) las deficiencias metodológicas y sus consecuencias, y f) los sesgos o conflictos de interés (reales o posibles) que puedan o no influir en la metodología, análisis, o interpretación y presentación de los descubrimientos.

Desarrolle capacidades de EI en organizaciones de profesionales. Cualquier serie de recomendaciones también debe tomar en cuenta explícitamente los inconvenientes del enfoque de prueba para la evaluación de impacto, entre los principales está el costo, la duración, la dificultad, y los requisitos técnicos, de los cuales todos han limitado su utilidad en cuanto a la gestión de programas. Debido a las restricciones técnicas y de recursos de las organizaciones de profesionales, un enfoque de EI que pone a la validez científica como el criterio principal, probablemente no sea adoptado por organizaciones de profesionales. Por otra parte, un enfoque en EI se centra en la utilidad de EI para la gestión de programas, y que puede ofrecer garantías razonables a la administración de que sus beneficios en términos de eficacia mejorada en el programa superan sus costos, probablemente sea adoptada por las organizaciones de profesionales a una escala más amplia. Este hecho debe ser reconocido desde el principio.

De modo que si el objetivo principal de EI es comprobar el impacto, y las autoridades están satisfechas con evaluar esporádicamente un número relativamente pequeño de programas, entonces la dependencia en EI científico probablemente sea adecuada. Pero si un objetivo de EI es mejorar el impacto, y las autoridades quieren información más o menos regular sobre el desempeño de los programas a través de una variedad más grande de programas, entonces la dependencia en EI científico probablemente no sea apropiada. En el último caso, la inversión pública en EI debe concentrarse cada vez más en desarrollar la capacidad de EI y de investigación de mercado de las organizaciones de profesionales, y asistir en el desarrollo e implementación de metodologías firmes, prácticas de bajo costo, accesibles para los profesionales, y útiles. Esto requiere a su vez que las autoridades colaboren con organizaciones de profesionales para comprender cuáles son sus restricciones y necesidades y cooperar con ellos para crear una política de fondos y ayuda técnica que transmita conocimientos y destrezas críticas para que los programas puedan efectuar sus propias evaluaciones de modo que sea útil para ellos y sus clientes. Además, para las autoridades que desean promover la sostenibilidad y continuidad, mejorar el impacto a través de EI e investigaciones de mercado

ofrece una de las propuestas para implementar políticas con más posibilidades de ser eficaces.

Notas

1. Una alta participación de los pobres en el sector de actividad informal es particularmente característica de los LDCs.
2. Para otras reseñas de metodologías de EI, véanse Gaile y Foster (1996), Khandker (1998), Hulme (2000), Coleman (2001a) y Schreiner (2002).
3. Quizás la mejor fuente para comprender los fundamentos conceptuales de la Evaluación del impacto y sus muchas relaciones complejas es la serie de artículos de investigación conceptuales comisionados y publicados por el Assessing the Impact of Microenterprise Services Project (AIMS) [Proyecto de Servicios de Evaluación del Impacto de la Microempresa], fundada por la Oficina para el Desarrollo Microempresarial en la USAID. Las publicaciones de la AIMS abarcan temas tales como evaluar el impacto al nivel de la empresa y del hogar (Inserrra, 1996) y a nivel individual (Chen, 1997), del ingreso y los activos como indicadores del impacto (Barnes, 1996, Little, 1997), la medición de las ganancias y el valor neto de las microempresas (Daniels, 1999), evaluar los impactos dentro de un “portafolio económico del hogar” (Chen y Dunn, 1996), y las microfinanzas y la administración del riesgo (Cohen y Sebstad, 1999). Todas las publicaciones del AIMS pueden ser descargadas de www.usaidmicro.org.
4. Los términos “Método científico”, “tradición humanista” y aprendizaje participativo son usados aquí para describir los enfoques metodológicos de la evaluación de impacto, también como resúmenes de cada enfoque, están basados en la excelente discusión de Hulme (2000) sobre metodologías para la evaluación de impacto.
5. Muchos de los estudios “participativos” son participativos solo de nombre. El término participación rápidamente alcanza el estatus de cliché- frecuentemente fuera de uso y de significado sustantivo.
6. SEEP (por sus siglas en inglés) se refiere a la Red de Educación y Promoción de las Pequeñas Empresas. SEEP es una red de profesionistas de organizaciones del voluntariado privado de Estados Unidos encargadas de promover las microfinanzas. SEEP trabaja conjuntamente con el equipo de investigación de AIMS para desarrollar las herramientas SEEP/AIMS.
7. Ver Nelson *et al.* (2001) para una descripción a profundidad de las instrucciones para implementar cada una de las herramientas SEEP/AIMS.
8. MicroSave Africa, por ejemplo ha desarrollado un conjunto de herramientas de investigación de mercado que se utilizan ampliamente en las metodologías PLA. A la fecha, MicroSave ha dirigido diversos grupos de capacitación alrededor del mundo que involucran a docenas de instituciones de microfinanzas. La información sobre las herramientas de investigación de mercado de MicroSave pueden ser revisadas en la página www.microsave-africa.com
9. Por ejemplo, un carpintero que pudo haberse mudado recientemente a la comunidad, o haber empezado su negocio, no debió haber escuchado del programa sino hasta recientemente.
10. Un ejemplo de cómo no se debe seleccionar un grupo de control está en Kosanovich y Fleck (2002) quienes eligieron a los miembros del grupo de control entre individuos a los que se les ofreció integrarse al programa microempresarial pero que desertaron. Este enfoque virtualmente garantiza un significativo sesgo de selección.

11. El enfoque de Pitt y Khandker fue criticado por Morduch (1999), quien cuestionó si los tres programas evaluaban en realidad forzados por las reglas de elegibilidad, y encontró otros problemas en la metodología econométrica de Pitt y Khandker. Pitt (1999) en respuesta a Morduch defendió el enfoque original y se propuso demostrar cómo era superior a la alternativa metodológica propuesta por Morduch.
12. Hulme (2000) señala que la división de los préstamos al consumo pueden de hecho producir mayores tasas de retorno que las inversiones de las empresas, por ejemplo, si el gasto del consumo toma la forma de inversión en capital humano (p. ej., becas escolares, cuidado de la salud) reemplaza el préstamo de otras fuentes con costos más altos, o si se utiliza para satisfacer necesidades básicas (p. ej., comida o medicina) necesariamente implica niveles sostenibles de productividad laboral.
13. También sería útil incluir clientes de tres años en el grupo de tratamiento, sin embargo, en muchas de las instituciones de microfinanciamiento, los clientes de tres o cuatro años son difíciles de conseguir, manteniendo tasas marginales de clientes altos.
14. Este autor, por ejemplo participó en evaluaciones de impacto longitudinales para programas de microfinanzas en Tanzania que sufría de más del 80 por ciento de desgaste del panel después de uno o dos años de estudio.
15. Soller y Parsons (2002), por ejemplo encontró que los programas de microfinanciamiento pueden contribuir a generar cientos o miles de millones de dólares para las economías locales a través de gastos directos y multiplicadores de ingreso. Otros impactos examinados a nivel de la comunidad local incluyen Coleman (1999, 2001b), Khandker (1996, 2001), Khandker et al. (1998), Kilkby y D'Zamura (1985), Morduch, (1998), Pitt et al. (1999), Schuler and Hashemi (1994), Schuler, Hashemi, and Riley (1997), y Zeller et al. (1996).
16. El costo de oportunidad del empleo ocurre cuando nuevos empleados dejan un trabajo previo y no son reemplazados o son reemplazados por trabajadores menos productivos.
17. Para una buena crítica que demuestra como las tasas de pago de los préstamos pueden ser engañosas véase Rosenberg (1999).

Apéndice

Recomendaciones para evaluaciones de alcance medio

En abril de 1997 y abril de 1998, el Grupo Consultivo para Ayudar a los Más Pobres (CGAP, por sus siglas en inglés) llevó a cabo dos reuniones virtuales de expertos en microfinanciamiento para crear pautas metodológicas para evaluaciones de alcance medio (Cohen and Gail, 1998). Llegaron a un consenso con las siguientes pautas:

1. Usar algún tipo de perspectiva de tiempo. Permita que transcurra suficiente tiempo para que ocurran los impactos (ambos en términos de participación de clientes y madurez del programa).
2. Use algún tipo de grupo de comparación. Los no clientes son preferibles a los clientes pivote dentro de lo posible. Sin embargo, el costo y limitaciones de enfoque más altos son bien reconocidos.
3. Cree evaluaciones a la medida del contexto específico que está siendo estudiado.
4. Empiece con un grupo pequeño de indicadores que hayan demostrado validez en estudios previos de EI y que sean relativamente fáciles de reunir. Incorpore nuevos indicadores a medida que sea adecuado.
5. Reúna indicadores de base, de ser posible cuando los clientes entren al programa. Si no es posible, use información retrospectiva.
6. Use información a nivel de intervalos cuando sea posible.
7. Use mejor EI como una herramienta de gestión para generar información que sea útil para la mejora del programa; por ejemplo, incorpore la satisfacción del cliente en los estudios de EI.

8. Emplee métodos para establecer una asociación posible entre la participación en el programa y los resultados observados.
9. Incorpore la satisfacción del cliente como parte del EI.
10. Use un EI diseñado cuidadosamente para asegurar transparencia y revisión o supervisión externas.
11. Incorpore planes para EI en el diseño del programa y la implementación tan pronto como sea posible.
12. Mida la dirección del cambio cuando no se pueda calcular el cambio.
13. Construya una capacidad local para realizar EI, de forma interna y externa a la organización.

Referencias bibliográficas

- ASHE, Jeffrey y Lisa PARROTT (2001), PACT's Women's Empowerment Program in Nepal.
- A Savings y Literacy Led Alternative to Financial Institution Building. Washington, DC: PACT.
- ASHE, Jeffrey y Michael MACINTYRE (2000), Working Capital. Building the Grass Roots Economy of Low Income Communities. Part 1: Membership Characteristics and Program Impact. Cambridge, MA: Working Capital.
- BARNES, Carolyn (1996), "Assets and the Impact of Microenterprise Finance Programs". AIMS Paper. Washington, DC: Management Systems International.
- BARNES, Carolyn (2001), "Microfinance Program Clients and Impact: An Assessment of Zambuko Trust, Zimbabwe". AIMS Paper. Washington, DC: Management Systems International.
- BARNES, Carolyn, Gaile MORRIS and Gary GAILE (1999), "An Assessment of Clients of Microfinance Programs in Uganda". *International Journal of Economic Development*, vol. 1, núm. 1, 80-121.
- BENDICK, M. y M.L. EGAN (1987), "Transfer Payment Diversion for Small Business Development: British and French Experience". *Industrial and Labour Relations Review*, vol. 40, núm. 4, 528-542.
- BENUS, Jacob M., Michelle WOOD y Neelima GROVER (1994), A Comparative Analysis of the Washington and Massachusetts UI Self-Employment Demonstrations. Bethesda, MD: Abt Associates.
- BLAIR, Amy Kay y Joyce KLEIN (2001), Microenterprise as a Welfare to Work Strategy: Client Characteristics. Washington, DC: The Aspen Institute.
- BOLNICK, Bruce R. y Eric R. NELSON (1990), "Evaluating the Economic Impact of a Special Credit Programme: KIK/KMKP in Indonesia". *The Journal of Development Studies*, vol. 26, núm. 2, 299-312.
- BUCKLEY, Graeme (1996a), "Financing the Jua Kali Sector in Kenya. The K-Rep Juhudi Scheme and Kenya Industrial Estates Informal Sector Program". In *Finance Against Policy*, volume II: Country Case Studies, ed. David Hulme and Paul Mosley, 271-332, Londres: Routledge.
- BUCKLEY, Graeme (1996b), "Rural and Agricultural Credit in Malawi. A Study of the Malawi Mudzi Fund and the Smallholder Agricultural Administration". In *Finance Against Policy*, Volume II: Country Case Studies, edited by David Hulme and Paul Mosley, 333-408, Londres: Routledge.
- BUEINIC, Mayra, Marguerite BERGER y Cecelia JARAMILLO (1989), "Impact of a Credit Project for Women and Men Microentrepreneurs in Quito, Ecuador". In *Women's Ventures*, ed. M. Berger and M. Buvinic, 222-246, West Hartford, CT: Kumarian Press.
- CHAMBERS, Robert (1997), *Whose Reality Counts? Putting the First Last*. London: IT Publications.
- CHEN, Martha Alter (1997), "A Guide for Assessing the Impact of Microenterprise Services at the Individual Level". AIMS Paper. Washington, DC: Management Systems International.

- CHEN, Martha Alter y Elizabeth DUNN (1996), "Household Economic Portfolios". AIMS Paper. Washington, DC: Management Systems International.
- CHEN, Martha Alter y Donald SNODGRASS (2001), "Managing Resources, Activities, and Risk in Urban India: The Impact of SEWA Bank". AIMS Paper. Washington, DC: Management Systems International.
- CHURCHILL, Craig F. (1995), "The Impact of Credit on Informal Sector Enterprises in South Africa: A Study of Get Ahead Foundation's Stokvel Lending Programme". MA Thesis, Worcester, MA: Clark University.
- CLARK, Peggy y Tracy HUSTON (1993), *Assisting the Smallest Businesses: Assessing Microenterprise Development as a Strategy for Boosting Poor Communities*. Washington, DC: The Aspen Institute.
- CLARK, Peggy, Amy KAYS, Lily ZANDNIAPOUR, Enrique SOTO y Karen DOYLE (1999), Washington DC: The Aspen Institute.
- COHEN, Monique (1999), "Opening Up the Impact Assessment Agenda". AIMS Paper. Washington, DC: Management Systems International.
- COHEN, Monique y Gary GAILE (1998), "Highlights and Recommendations of the Second Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies: April 14-28, 1998. Developing Lower Cost Microenterprise Impact Assessment Methodologies for Microenterprise Programs". AIMS Paper. Washington, DC: Management Systems International.
- COHEN, Monique y Jennefer SEBSTAD (1999), "Microfinance and Risk Management: A Client Perspective". AIMS Paper. Washington, DC: Management Systems International.
- COHEN, Monique y Jennefer SEBSTAD (2000), "Microfinance, Risk Management, and Poverty". AIMS Paper. Washington, DC: Management Systems International.
- COLEMAN, Brett E. (1999), "The Impact of Group Lending in Northeast Thailand". *Journal of Development Economics*, vol. 60, núm. 2, 105-142.
- COLEMAN, Brett E. (2001a), "Measuring Impact of Microfinance Programs". *ADB Finance for the Poor*, vol. 2, núm. 4, 5-8
- COLEMAN, Brett E. (2001b), "Microfinance in Northeast Thailand: Who Benefits and How Much?" Mimeo. Asian Development Bank.
- COPESTAKE, James (2000), "Integrating Impact Monitoring and Assessment of Microfinance". *Journal of Development in Practice*, vol. 10, núm. 5, 705-711.
- COPESTAKE, James, Sonia BHALOTRA y Susan JOHNSON (2001), "Assessing the Impact of Microcredit: A Zambian Case Study". *The Journal of Development Studies*, vol. 37, núm. 4, 81-100.
- CREEVEY, Lucy E. (1994), "Summary Report: A UNIFEM Comparative Study Assessing Impacts of Projects to Support Microenterprises".
- CREEVEY, Lucy E., Koumakh NDOUR y Abdourahmane THIAM (1995), "Evaluation of the Impacts of PRIDE/EITA (The Guinea Rural Enterprise Development Project)". GEMINI Technical Report, núm. 94. Bethesda, MD: Development Alternatives, Inc.
- DANIELS, Lisa (1999), "Alternatives for Measuring Profits and New Worth of Microenterprises". 1999. AIMS Paper. Washington, DC: Management Systems International.

- DEARDEN, Kirk y Nazmul KHAN (1994), "Assessing the Impact of Women's Savings and Credit Programs on Fertility: A Case from Bangladesh". *Save the Children Women/Child Impact Program Monograph*, núm. 2, Wesport: CT.
- DIAGNE, Aliou (1998), "Impact of Access to Credit on Income and Food Security in Malawi". FCND Discussion Paper núm. 46. Washington, DC: International Food Policy Research Institute.
- DUNN, Elizabeth y Arbuckle J. GORDON Jr. (2001), "The Impact of Microcredit: A Case Study from Peru". AIMS Paper. Washington, DC: Management Systems International.
- DRURY, David, Stephen WALSH y Marlene STRONG (1994), Evaluation of the EDWAA Job Creation Demonstration. Washington, DC: US Department of Labour, Employment and Training Administration.
- DUMAS, Colette (2001), "Evaluating the Outcomes of Microenterprise Training for Low Income Women: A Case Study". *Journal of Developmental Entrepreneurship*, vol. 6, No.2, 97-128.
- ELSE, John F. y Carmel CLAY-THOMPSON, Refugee Microenterprise Development: Achievements and Lessons Learned. Iowa City: Institute for Social and Economic Development.
- GAILE, Gary L. y Jennifer FOSTER (1996), "Review of Methodological Approaches to the Study of the Impact of Microenterprise Credit Programs". AIMS Paper. Washington, DC: Management Systems International.
- GOETZ, Anne Marie y Rina SEN GUPTA (1996), "Who Takes the Credit? Gender, Power, and Control Over Loan Use in Rural Credit Programs in Bangladesh". *World Development*, vol. 24, núm. 1, 45-63.
- GUPTA, Surendra y Mario D. DAVALOS (1993), "Midterm Evaluation of the Microenterprise Development Project of Jamaica". GEMINI Technical Report núm. 59. Bethesda, MD: Development Alternatives, Inc.
- HASHEMI, Syed M., Sidney Ruth SCHULER y Ann P. RILEY (1996), "Rural Credit Programs and Women's Empowerment in Bangladesh". *World Development*, vol. 24, núm. 4, 635-53.
- HIMES, Cristina y Lisa J. SERVON (1998), Measuring Client Success: An Evaluation of ACCION's Impact on Microenterprises in the United States. Cambridge, MA: ACCION International.
- HULME, David (2000), Impact Assessment Methodologies for Microfinance: Theory, Experience, y Better Practice. *World Development*, vol. 28, núm. 1, 79-98.
- HULME, David, Richard MONTGOMERY, con Debapriya BHATTACHARYA (1996), "Mutual Finance and the Poor. A Study of the Federation of Thrift and Credit Cooperatives in Sri Lanka (SANASA)". In *Finance Against Policy*, volume II: Country Case Studies, ed. David Hulme and Paul Mosley, 177-245, Londres: Routledge.
- INSERRA, Anne (1996), "A Review of Approaches for Measurement of Microenterprise and Household Income. AIMS Paper". Washington, DC: Management Systems International.
- INSTITUTE FOR SOCIAL AND ECONOMIC DEVELOPMENT (1994), "Rivercities of Iowa/Illinois Self-Employment (RISE) Final Report". Iowa City: Institute for Social and Economic Development.

- KARLAN, Dean y Gwendolyn ALEXANDER (2002), "Credit to the Poor: Examining the Attrition Bias of Cross-Sectional Impact Assessments". Mimeo. Princeton University.
- KEVANE, Michael y Bruce WYDICK (2001), "Microenterprise Lending to Female Entrepreneurs: Sacrificing Economic Growth for Poverty Alleviation". *World Development*, vol. 29, núm. 7, 1225-1236.
- KHANDKER, Shahidur R. (1996), "Grameen Bank: Impact, Costs, and Program Sustainability". *Asian Development Review*, vol. 14, núm. 1, 97-130.
- KHANDKER, Shahidur R. (1998), "Micro-Credit Programme Evaluation: A Critical Review". *IDS Bulletin*, vol. 29, núm. 4, 11-20.
- KHANDKER, Shahid (2001), "Does Micro-Finance Really Benefit the Poor: Evidence from Bangladesh". Paper presented at the Asia and Pacific Forum on Poverty: Reforming Policies and Institutions for Poverty Reduction, Manila, febrero 5-9.
- KHANDKER, Shahidur R., Hussain A. SAMAD y Zahed H. KHAN (1998), "Income and Employment Effects of Micro-Credit Programmes: Village-Level Evidence from Bangladesh". *Journal of Development Studies*, vol. 35, núm. 2, 96-124.
- KILBY, Peter y David D'ZMURA (1985), "Searching for Benefits". USAID Special Study núm. 28. Washington, DC: United States Agency for International Development.
- KOSANOEICH, William T. y Heather FLECK (2002), *Comprehensive Assessment of Self-Employment Assistance Programs*. Washington, DC: US Department of Labour.
- LAPAR, MA. Lucila A., Douglas H. GRAHAM y Richard L. MEYER (1995), "The Effect of Credit on Output: Are the Sectoral Differences?" Rural Finance Program. Department of Agricultural Economics. Columbus, OH: The Ohio State University.
- LAPAR, MA. Lucila A., Douglas H. GRAHAM, Richard L. MEYER y David S. KRAYBILL (1995), "Selectivity Bias in Estimating the Effect of Credit on Output: The Case of Rural Non-Farm Enterprises in the Philippines". Rural Finance Program. Department of Agricultural Economics. Columbus, OH: The Ohio State University.
- LITTLE, Peter D. (1997), "Income and Assets as Impact Indicators". AIMS Paper. Washington, DC: Management Systems International.
- MCKERNAN, Signe-Mary (1996), "The Impact of Microcredit Programmes on Self-Employment Profits: Do Non-credit Programme Aspects Matter?" *The Review of Economics and Statistics*, vol. 84, núm. 1, 93-115.
- MKNELLY, Barbara, Chatree WATETIP, Cheryl A. LASSEN y Christopher DUNFORD (1996), "Preliminary Evidence that Integrated and Financial and Educational Services Can Be Effective Against Hunger and Malnutrition". *Research paper* núm. 2. Davis, CA: Freedom from Hunger.
- MKNELLY, Barbara y Christopher DUNFORD (1999a), "Impact of Credit with Education on Mothers and their Young Children's Nutrition: CRECER Credit with Education Program in Bolivia". *Research Paper* núm. 5. Davis, CA: Freedom from Hunger.
- MKNELLY, Barbara y Christopher DUNFORD (1999b), "Impact of Credit with Education on Mothers and their Young Children's Nutrition: Lower Pra Rural Bank Credit with Education Program in Ghana". *Research Paper* núm. 4. Davis, CA: Freedom from Hunger.

- MONTGOMERY, Richard, Debapriya BHATTACHARYA y David HULME (1996), "Credit for the Poor in Bangladesh. The BRAC Rural Development Programme and the Government Thana Resource Development and Employment Programme". En *Finance Against Poverty*, volume II: Country Case Studies, ed. David Hulme and Paul Mosley, 94-176, Londres: Routledge.
- MORDUCH, Jonathan (1998), "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh". Cambridge, MA: Harvard University.
- MOSLEY, Paul (1996a), "India. The Regional Rural Banks". In *Finance Against Policy*, Volume II: Country Case Studies, ed. David Hulme y Paul Mosley, 246-270, Londres: Routledge.
- MOSLEY, Paul (1996b), "Indonesia: BKK, KUCK, and the BRI Unit Desa Institutions". En *Finance Against Poverty*, volume II: Country Case Studies, ed. David Hulme y Paul Mosley, 32-93, Londres: Routledge.
- MOSLEY, Paul (1996c), "Metamorphosis from NGO to Commercial Bank: The Case of Bancosol in Bolivia". En *Finance Against Policy*, volume II: Country Case Studies, ed. David Hulme y Paul Mosley, 1-31, Londres: Routledge.
- MOSLEY, Paul (2001), "Microfinance and Poverty in Bolivia". *The Journal of Development Studies*, vol. 37, núm. 4, 101-132.
- MOSLEY, Paul y David HULME (1998), "Microenterprise Finance: Is There a Conflict Between Growth and Poverty Alleviation?" *World Development*, vol. 26, núm. 5, 783-790.
- MT. AUBURN ASSOCIATES (1998), *Evaluation of Working Capital Delaware*. Somerville, MA: Mt. Auburn Associates.
- MUSTAFA, SHAMS, ISHRAT ARA, DILRUBA BANU, ALTAF HOSSAIN, AJMAL KUBIR, MOHAMMAD MOHSIN y ABU YUSUF (1995) "Impact Assessment Study of BRAC's Rural Development Programme". *Final Report*.
- NEILL, Catherine, Mario DAVALOS, Washington KIIRU, M. MANUNDU y Jennefer SEBSTAD (1994), "The Kenya Rural Enterprise Programme under Cooperative Agreement núm. AID-615-0282-A-00-7026-00: A Final Evaluation". GEMINI Technical Report núm. 77. Bethesda, MD: Development Alternatives, Inc.
- NELSON, Eric (1984), "The Economic Impact of KIK/KMKP Credit on Indonesian Small Enterprises". Draft Memo to Kapala Urusan Kredit Koperasi dan Kredit Kecil. Mimeo.
- NELSON, Eric y Bruce R. BOLNICK (1986), "Survey Methods for Assessing Small Credit Programs Evaluating the Economic Impact of KIK/KMPK Credits". Jakarta.
- NELSON, Candace, Barbara MKNELLY, Elaine EDGCOMB, Gary GAILE, Carter GARBER, Nancy HORN, Karren LIPPOLD y Brian BEARD (2001), "Learning from Clients: Assessment Tool for Microfinance Practitioners". AIMS Paper. Washington, DC: Management Systems International.
- OLDHAM, Linda, Hager EL HADIDID, Hussein TAMA, Michele SULIMAN NASHED, SHERIF EL DEWANY, Mahmoud HUSSEIN, Mohammed ABDEL AZIZ y Mohamed SAKR (1994), "Measuring Socioeconomic Impact of Credit on SMI: Assessment of the Monitoring System used by the Alexandria Businessmen's Association", Egypt. GEMINI Technical Report núm. 76. Bethesda, MD: Development Alternatives, Inc.

- PARK, Albert y Changqing REN (2001), "Microfinance with Chinese Characteristics". *World Development*, vol. 29, núm. 1, 39-62.
- PITT, Mark M. (1999), "Reply to Jonathan Morduch's 'Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh'", Providence, RI: Brown University.
- PITT, Mark M. y Shahidur R. KHANDKER (1996), "Household and Intrahousehold Impacts of the Grameen Bank and Similar Targeted Credit Program in Bangladesh". *World Bank Discussion Paper* núm. 320. Washington, DC: World Bank.
- PITT, Mark M. y Shahidur R. KHANDKER (1998), "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy*, vol. 106, núm. 5, 958-996.
- PITT, Mark M., Shahidur KHANDKER, Osman H. CHOWDHURY y Daniel L. MILLIMET (1999), "Credit Programs for the Poor and Health Status of Children in Rural Bangladesh". *PSTC Working Paper* núm. 98-01. Providence, RI: Brown University.
- PITT, Mark M., Shahidur KHANDKER, Signe-Mary MCKERNAN y M. ABDUL LATIF (1999), "Credit Programs for the Poor and Reproductive Behaviour in Low-Income Countries: Are Reported Causal Relationships the Result of Heterogeneity Bias?" *Demography*, vol. 36, núm. 1, 1-21.
- PULLEY, Robert V. (1989), "Making the Poor Creditworthy". A Case Study of the Integrated Rural Development Program in India. Washington, DC: The World Bank.
- RAHEIM, Salome (1996), "Microenterprise as an Approach for Promoting Economic Development in Social Work: Lessons from Self-Employment Investment Demonstration". *International Journal of Social Work*, vol. 39, 69-82.
- RAHEIM, Salome y Jason FRIEDMAN (1999), "Microenterprise in the Heartland: Self-Employment as a Self-Sufficiency Strategy for TANF Recipients in Iowa 1993-1998". *Journal of Microfinance*, vol. 1, núm. 1, 66-90.
- ROSENBERG, Richard (1999), "Measuring Microcredit Delinquency: Ratios Can Be Harmful to Your Health". *Occasional Paper* núm. 3. Washington, DC: Consultative Group to Assist the Poorest.
- SCHREINER, Mark (2002), "Evaluation and Microenterprise Programs in the United States". *Journal of Microfinance*, vol. 4, núm. 2, 67-91.
- SCHULER, Sidney Ruth y Syed M. HASHEMI (1994), "Credit Programs, Women's Empowerment, and Contraceptive Use in Rural Bangladesh". *Studies in Family Planning*, vol. 25, núm. 2, 65-76.
- SCHULER, Sidney Ruth, Syed M. HASHEMI y Ann P. RILEY (1997), "The Influence of Women's Changing Roles and Status in Bangladesh's Fertility Transition: Evidence from A Study of Credit Programs and Contraceptive Use". *World Development*, vol. 25, núm. 4, 563-575.
- SCHULER, Sidney Ruth, Syed M. HASHEMI y Shamsul HUDA BADAL (1998), "Men's Violence Against Women in Rural Bangladesh: Undermined or Exacerbated by Microcredit Programmes?" *Development in Practice*, vol. 8, núm. 2, 148-157.
- SEBSTAD, Jennefer (1992), "Get Ahead Foundation Credit Programs in South Africa: The Effects of Loans on Client Enterprises". *GEMINI Technical Report*, núm. 44. Bethesda, MD: Development Alternatives, Inc.

- SEBSTAD, Jennefer y Sara LOZA (1993), "Lending and Learning: Formal Banks and Microenterprise in Egypt". Reporte para la Fundación Ford, Cairo. Washington, DC: Community Economics Corporation.
- SEBSTAD, Jennefer y Martin WALSH (1991), "Microenterprise Credit and its Effects in Kenya: An Exploratory Study". Report prepared for USAID AFR/MDI and SandT/WID. Washington, DC: Coopers and Lybrand.
- SEKKESAETER, Unni Beate (2002), Evaluation of Hordaland Network Credit. A Microfinance Programme in Hordaland County, Norway — Part of the "Equal Credit" Project under the EU Receipte II Programme. Bradford, West Yorkshire: Bradford School of Management.
- SMITH, Stephen C. (2002), "Village Banking and Maternal and Child Health: Theory and Evidence from Ecuador and Honduras". World Development, vol. 30, núm. 4, 707-723.
- STEELE, Fiona, Sajeda AMIN y Ruchira T. NAVED (2001), "Savings/Credit Group Formation and Change in Contraception". Demography, vol. 38, núm. 2, 267-282.
- SUTORO, Ann Dunham (1990), "KUPEDS Development Impact Survey: Briefing Booklet". BRI, Planning, Research and Development Department.
- THE ROBERTS FOUNDATION (1995), Self Employment and Very Low-Income Women. A Final Report on the Experience of the San Francisco Homeless Women's Economic Development Project. San Francisco: The Roberts Foundation.
- TODD, Helen (1996), Women at the Center. Grameen Bank Borrowers after One Decade. Boulder: Westview Press.
- US DEPARTMENT OF HEALTH AND HUMAN SEREICES (1994a), "Capital Opportunities Expansion Program. Human Resource Development Council of District IX, Inc.: Bozeman, Montana". En Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990, 54-83, Washington, DC: US Department of Health and Human Services.
- US DEPARTMENT OF HEALTH AND HUMAN SEREICES (1994b), "Micro-Business Development Program. Central Vermont Community Action Council, Inc.: Barre, Vermont". En Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990, 1-19, Washington, DC: US Department of Health and Human Services.
- US DEPARTMENT OF HEALTH AND HUMAN SEREICES (1994c), "Microenterprise Development Program (MEDP). Mayor's Office of Community Services: Philadelphia, Pennsylvania". En Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990, 38-53, Washington, DC: US Department of Health and Human Services.
- US DEPARTMENT OF HEALTH AND HUMAN SEREICES (1994d), "Small Enterprise and Family Development. Southeast Iowa Community Action: Burlington, Iowa". En Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990, 20-37, Washington, DC: US Department of Health and Human Services.
- VENGROFF, Richard y Lucy CREEVEY (1994), Evaluation of Project Impact: ACEP Component of the Community Enterprise Development Project. Storrs: The University of Connecticut.

- WOLLER, Gary y Robert PARSONS (2002), "Assessing the Community Economic Impact of Microfinance Institutions", *Journal of Developmental Entrepreneurship*, vol. 7, núm. 2, 133-150.
- WYDICK, W. Bruce (1999a), "Credit Access, Human Capital, and Class Structure Mobility". *The Journal of Development Studies*, vol. 55, núm. 6, 131-152.
- WYDICK, W. Bruce (1999b), "The Effect of Microenterprise Lending on Child Schooling in Guatemala". *Economic Development and Cultural Change*, vol. 47, núm. 4, 853-869.
- ZAMAN, Hassan (2001), "Assessing the Poverty and Vulnerability Impact of Micro-Credit in Bangladesh: A Case Study of BRAC". Washington, DC: Consultative Group to Assist the Poorest.
- ZELLER, Manfred, Akhter AHMED, Suresh BABU, Sumiter BROCA, Aliou DIAGNE y Manohar SHARMA (1996), "Rural Financial Policies for Food Security of the Poor. Methodologies for a Multicountry Research Project". Washington, DC: International Food Policy Research Institute.

Capítulo 18

Un panorama del panel de discusión: evaluando el desarrollo económico local y el empleo

Por
Alice Nakamura,
Facultad de Negocios, Universidad de Alberta,
Edmonton, Alberta, Canadá

Alistair Nolan del Secretariado de la OCDE/LEED presentó el último panel de debate. La pregunta inicial que planteó fue: ¿cuál es el estado actual del compromiso gubernamental hacia la evaluación? ¿Es necesario mejorar la situación actual? Si ese es el caso, ¿cómo? Nolan dijo que el debate debería enfocarse, en primer lugar, en los problemas prácticos de política asociados con los proyectos de evaluación. Él comentó que a los panelistas se les había solicitado que hicieran un comentario inicial, y que posteriormente se abriría la sesión para preguntas. El debate subsiguiente permitió ilustrar la importancia de diversos aspectos técnicos y de política planteados en las presentaciones de otras conferencias.

Edward W. Hill, profesor de la Universidad del Estado de Cleveland en Estados Unidos, comenzó esbozando lo que él veía como temas clave para el debate:

- La definición de una política de desarrollo económico local y sus resultados.
- Las diferencias entre procesos y evaluaciones acumulativas y los modos en los cuales cada una de ellas pueden ser útiles para la estructura causal de los programas, tanto los que se encuentran en ejecución como los que se están ideando.

Los problemas asociados con el hecho de que las evaluaciones locales no se realizan a ese nivel.

Hill señala que las políticas para el empleo y el desarrollo económico son muy diferentes. Llama la atención hacia el reto de distinguir las políticas laborales de la inversión laboral, y las actividades para el gasto en el desarrollo de la inversión financiera en un contexto comunitario. Considera que se debería poner más énfasis al hecho de saber si las políticas impactan en el lado de la demanda o en el de la oferta de la economía. Argumenta también en contra de tratar a la política laboral como un simple instrumento del desarrollo económico. Desde una perspectiva de desarrollo económico, señala la rentabilidad de la empresa como una medida problemática que en primer lugar refleja consideraciones de rendición de cuentas y sobre el código fiscal nacional. Cita el desastre de Enron como indicador de algunos de los defectos de enfocarse en la rentabilidad de la empresa. Sugirió que las tasas de supervivencia de las empresas y los indicadores sobre innovación de producto, que toman en cuenta el ciclo de vida del producto,

pueden proveer una mejor guía para los indicadores de producción y podrían ser más útiles para los propósitos de desarrollo económico. También argumenta que el ingreso y las tasas de crecimiento en el empleo deberían tener mayor prominencia en la planeación del desarrollo económico local debido a que reflejan el equilibrio entre la demanda y la oferta de los mercados.

Henry B. McLeish, miembro del parlamento escocés y anterior primer ministro de Escocia, fue el siguiente en hablar. Indicó que veía una erosión en la confianza hacia la evaluación y el desarrollo económico. Sentía también esa confusión que se ha desarrollado respecto a los resultados de la evaluación y sus procesos. Instó a tener una mayor precaución al hacer el proceso de transmisión del conocimiento simple y transparente. Exhortó a aquellos que se embarcan en proyectos pilotos a tener en mente tres objetivos. El primero consiste en utilizar la información obtenida de evaluaciones anteriores para reformular los proyectos de evaluación. El segundo busca tener cuidado al pensar en cómo medir el éxito al finalizar un proyecto. Y el tercero trata de encontrar mecanismos para aplicar proyectos de desarrollo tangencial para mejorar otros aspectos de la política pública y los procesos más allá de los objetivos previstos del proyecto.

Para finalizar, McLeish señaló que los diseñadores de políticas tienen una inevitable tendencia a tratar de guardarse los detalles del proceso de evaluación, y que este tipo de interferencia política generalmente no es tan transparente como sería deseable. Asimismo, dijo que el aprendizaje (capital humano) es un factor crucial en el desarrollo económico. Dijo también que la calidad de vida es una dimensión que no ha sido reconocida adecuadamente, y argumentó que la calidad de vida no está correctamente reflejada en la mayoría de los indicadores sobre desarrollo económico que actualmente se utilizan.

El siguiente panelista en hablar fue Stephen Wandner del Departamento del Trabajo de Estados Unidos. Wandner indicó que resaltaría las características distinguibles clave para los programas de capacitación y empleo en Estados Unidos. Dijo que una fortaleza característica de los programas públicos en Estados Unidos es que la evaluación es un componente integral y obligatorio. Citó el ejemplo del Decreto para Inversión en el Lenguaje y en la Fuerza Laboral de 1998, el cual recomendaba la implementación de un proceso de evaluación que valorara correctamente el grado de éxito conseguido con la realización de los objetivos centrales del Decreto. Señaló que este proceso de evaluación usaba grupos de control y el método científico de la asignación aleatoria. Explicó que las evaluaciones modernas de este tipo en Estados Unidos constituyen un proceso abierto. Comentó que esa apertura incluía desde darle al público el derecho de ver todo a partir de los datos recolectados (después de que se tomaron las medidas convenientes para garantizar la privacidad de los individuos del proyecto) hasta los reportes de evaluación preliminares y los estudios finales. Esos materiales están disponibles al público, incluyendo, por supuesto, a la

comunidad universitaria de investigación. Reconoció que, inevitablemente, no a todos los que están dentro de las comunidades de investigación y de políticas les agradan todos los aspectos de cómo se condujeron las evaluaciones o las conclusiones que se derivaron de los reportes de resultados. Sin embargo, dijo que en Estados Unidos se acepta que las evaluaciones dirigidas correctamente y la apertura para acceder a la información generada por ellas es esencial para el debate informado sobre aspectos que enfrenta la sociedad y que, en la era de la información, la apertura de este tipo es un legado y un aspecto esencial de la fortaleza del sistema político estadounidense.

Alice Nakamura, profesora de la Escuela de Negocios de la Universidad de Alberta, Canadá, fue la siguiente en hablar. Respaldo ampliamente los comentarios de Wandner acerca de la importancia de la evaluación y el acceso abierto a las bases de datos y resultados derivados de la evaluación. Indicó que en Canadá se tienen también algunas experiencias muy favorables sobre este enfoque; comentó que Canadá fue un país pionero en la evaluación a finales de los ochenta y principios de los noventa con la experiencia del departamento gubernamental para el Desarrollo de los Recursos Humanos de Canadá (DRHC), responsable del programa de seguridad social para trabajadores desempleados (originalmente Seguro de Desempleo o SD, y ahora, Seguro para el Empleo, o SE, desde que se aprobó la iniciativa de Ley C-12 en 1996). Dijo que el DRHC había creado el camino, en asociación con el área de Estadísticas en Canadá, para desarrollar bases de datos para evaluar los programas SD/SE y establecer mecanismos que den acceso a las bases de datos a los investigadores ajenos al gobierno.

Nakamura dijo que la DRHC ha generado un liderazgo intelectual y de apoyo financiero con la formación de una amplia red, el Foro Canadiense de Investigación para el Empleo (FCIE), que consiste en una base de investigadores universitarios y del gobierno con habilidades metodológicas en áreas de evaluación de programas y análisis de empleos y salarios. La investigadora comentó que la iniciativa de la DRHC ha sido exitosa en redirigir la atención de los académicos encargados de realizar investigaciones hacia la evaluación de los programas canadienses. Antes de la creación del FCIE, los investigadores académicos canadienses interesados en la evaluación de los programas trabajaban, principalmente, con datos de Estados Unidos. Comentó que además de mejorar la calidad de la información para los programas de desarrollo en Canadá los datos de la DRHC y las iniciativas de investigación de la red, mejoraron ampliamente la calidad de la información sustantiva acerca de los programas SD/SE que los profesores habían difundido en los salones de clase universitarios de Canadá. De hecho, la iniciativa de DRHC ofrecía tangencialmente los beneficios sustanciales de los que hablaba McLeish.

El último panelista en hablar fue el profesor Philip Davies de la Oficina del Gabinete en el Reino Unido. Argumentó que en el Reino Unido ha habido una especie de renacimiento de la “política basada en la evidencia”. Dijo que diversos

reportes recientes han demostrado un incremento en los ajustes institucionales y en el financiamiento de las evaluaciones de política. Davis mencionó también que en los últimos cinco años ha habido una enorme proliferación en la investigación y en la industria de la evaluación. Comenzó entonces a recomendar diversas mejoras que consideraba, deberían hacerse al divulgar la información de los proyectos de evaluación en los contextos no académicos. Sugirió a los empleados que para los siguientes diseños y métodos de evaluación se aprovechen los datos que se han acumulado en los proyectos de evaluación anteriores para disminuir las presiones de tiempo, que considera inevitables, en los proyectos financiados por el gobierno. También sugirió asignar estratégicamente el financiamiento para la investigación. Concluyó su intervención señalando tres tipos de integración que considera necesarios para dar valor a las evaluaciones y que deben ser maximizados por los diseñadores de políticas. El primero consiste en la integración de las evaluaciones de política en el sentido de observar los resultados obtenidos en el ámbito económico, social, de distribución y riesgo. El segundo consiste en la integración de las metodologías incluyendo aquellas evaluaciones multinivel que incorporan datos nacionales y supranacionales. Y el tercero es la integración en el diseño e implementación de la política en un contexto de evaluación.

Siguiendo los comentarios iniciales de los panelistas, la discusión se abrió a las preguntas tanto de Nolan como de la audiencia. La ronda inicial de preguntas versó sobre los siguientes temas acerca del papel de las autoridades gubernamentales a nivel central:

- ¿Hacia dónde está dirigiéndose el impulso para mejorar la evaluación a nivel local?
- ¿Debe provenir del centro?
- ¿Debe involucrar la obligatoriedad de las evaluaciones y los incentivos correctos para impulsar evaluaciones de mayor calidad?

En respuesta, Wandner dio un ejemplo del Decreto de Inversión en Mano de Obra. Dijo que el Departamento del Trabajo se ha reunido con servidores públicos y los ha impulsado a construir capacidades para la evaluación. Sin embargo, admitió que las restricciones sobre las capacidades continuaban siendo un problema en algunos estados, incluyendo los más pequeños. McLeish dijo que sentía que al nivel local hay una cultura menos establecida de evaluación progresiva. Consideraba que la cultura en muchas localidades debería cambiar para enfocarse menos en los resultados de los proyectos y más en los procesos adaptados en dichos proyectos y las complejas innovaciones.

Hill argumenta que es importante tener presente las fuentes de financiamiento para las evaluaciones. Sugiere que, al nivel local, las evaluaciones acumulativas son menos importantes que las evaluaciones de proceso.

Nakamura sugirió que podría haber una interfaz potencial de negocios para las evaluaciones locales. Dijo que muchos negocios emprenden sus propias evaluaciones locales sobre una base regular, si bien reconoce que los negocios parecen estar más interesados en el proceso que en las evaluaciones acumulativas. Davis considera que al tener al mismo partido gobernante hace que le pertenezcan los problemas de desarrollo local y la evidencia de evaluación ayuda a alinear los incentivos para los evaluadores.

Al compartir historias sobre la evaluación se vuelve claro que ha habido un registro importante de aplicaciones exitosas en diversos países.

Finalmente, se puso atención en cómo se podría facilitar el acceso público a los datos de la evaluación y a los reportes y al mismo tiempo verificar que se protege la privacidad de los individuos. Parece claro que, en la era de la información, persistirán las tensiones actuales entre los instintos políticos y la obligación de un análisis del programa informado. También parece claro que, mientras que todo indica que Estados Unidos se ha inclinado por tener como prioridad el hacer público el acceso a los datos y reportes de la evaluación, hay muchos otros países en los que esto no ha ocurrido.

También hubo una discusión acerca de encontrar mecanismos para mejorar la comunicación de la información que genera la evaluación en los medios masivos. Esta discusión siguió con la rápida reflexión de Hill acerca de la importancia de enfocar a las diferentes audiencias de modos distintos. Hill añadió que los departamentos gubernamentales necesitan establecer estrategias de comunicación formal al principio de los proyectos de evaluación.

Acerca de los autores y colaboradores

Dr. Timothy Bartik

Timothy Bartik es economista en jefe en el Instituto W. E. Upjohn para la investigación sobre el empleo, una organización independiente sin fines de lucro y no partidista, ubicada en Kalamazoo, Michigan. Es doctor en economía por la Universidad de Wisconsin-Madison (1982). Fue profesor asistente de economía en la Universidad de Vanderbilt antes de unirse al instituto (1989), donde el doctor Bartik es responsable de dirigir las investigaciones sobre las políticas de desarrollo económico locales y estatales, sobre el mercado laboral local y sobre problemas de pobreza urbana. Ha escrito dos libros: *¿Quiénes se benefician de las políticas de desarrollo económico locales y estatales?* (Instituto Upjohn, 1991), y *Trabajos para los pobres: ¿pueden ayudar las políticas de demanda laboral?* (Fundación Russell Sage, 2001). Entre sus numerosos artículos académicos destacan: “Estrategias para el desarrollo económico”, en *Políticas de Gestión en las Finanzas Gubernamentales Locales* (en colaboración con J. R. Aronson y E. Schwartz); “Políticas para el desarrollo económico de Michigan” (con P. Eisinger y G. Erickcek) en *Michigan en el Milenio* (Universidad de Michigan, 2003); “Efectos multiplicadores de las reformas en la seguridad social en los mercados laborales estatales”, *Journal of Regional Science* (noviembre, 2002); “¿Pueden los programas de desarrollo económico ser evaluados?” (en colaboración con R. Bingham) en *Dilemas del desarrollo económico urbano* (Sage, 1997), y “¿Quiénes se benefician del crecimiento en el empleo local, los inmigrantes o los residentes originales?” (*Regional Studies*, septiembre, 1993).

W. E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007
USA
Tel.: +1 (616) 385 0433
Fax: +1 (616) 343 3308
Correo electrónico: BARTIK@we.upjohninst.org

* * *

Profesor Daniele Bondonio

Daniele Bondonio es profesor asistente en la Università del Piemonte Orientale, Alejandría, Italia. Imparte clases en las asignaturas de econometría y métodos cuantitativos para el programa de evaluación y análisis de políticas públicas en la Università del Piemonte Orientale y en la Università di Torino. Es doctor en análisis de política pública y gestión egresado de la Carnegie Mellon University, donde fue académico visitante en el 2001. Obtuvo su licenciatura en economía con mención honorífica por la Università di Torino. Los principales intereses de investigación del profesor Bondonio se enfocan en el impacto de las evaluaciones en el desarrollo económico local y en los programas de incentivos a empresas focalizados geográficamente. En su tesis de doctorado, analiza varios programas regionales de empresas estatales en Estados Unidos. Recibió la beca de investigación de la National Science Foundation para realizar su tesis doctoral y el premio a la tesis doctoral del Departamento para la Vivienda y Desarrollo Económico. Sus publicaciones recientes se enfocan en las evaluaciones de impacto de los programas de incentivos a empresas en la Unión Europea en áreas industrialmente deterioradas, de inversiones de subsidios focalizados hacia las empresas de reciente creación y los métodos para evaluar, comparativamente los programas de desarrollo económico implementados heterogéneamente entre diferentes regiones o estados. El profesor Bondonio también ha generado reportes de consultoría para los gobiernos regionales de Italia y forma parte del Proyecto de Evaluación en Turín, Italia, orientado a promover la práctica de evaluaciones de impacto en la administración pública.

Department of Public Policy and Public Choice
Università del Piemonte Orientale
Via Cavour 84
15100 Alessandria
Italia
Tel.: +39 011 533191/0131 283712
Fax: +39 011 5130721
Correo electrónico: daniele.bondonio@sp.unipmn.it

* * *

Dra. Paola Casavola

Dra. Paola Casavola recibió el título de leyes (1986) y el doctorado en economía (1991) por la Universidad de Nápoles, Italia. Es maestra en ciencias económicas por la London School of Economics (1989). De 1991 a 1999 trabajó como economista en el Departamento de Investigación (Servizio Studi) del Banco Central de Italia en Roma y realizó investigaciones principalmente

sobre el mercado laboral, el desarrollo local y la gobernabilidad corporativa. Desde mediados de 1999 ha trabajado en la Unidad de Evaluación (UVAL) del Departamento para el Desarrollo (DPS) del Ministerio Italiano de Economía (MEF). Dentro de la Unidad de Evaluación, es coordinadora del área de evaluación de programas y responsable de un conjunto de actividades dirigidas a construir y difundir herramientas metodológicas para la evaluación de los programas y para la administración de proyectos de evaluación. Dentro del Departamento para el Desarrollo es miembro del Comité de Coordinación del Reporte Anual sobre Políticas Territoriales. Ella también continúa su trabajo de investigación sobre el mercado laboral y las políticas de desarrollo local.

Dipartimento per le Politiche di Sviluppo e Coesione
 Unità di Valutazione
 Via Nerva 1
 00187 Rome
 Italia
 Tel.: +39 06 4761 9079
 +39 06 4761 9040 (secretaria)
 Fax: +39 06 4761 9037-47619075
 Correo electrónico: paola.casavola@tesoro.it

* * *

Dr. Philip Davies

Dr. Philip Davies actualmente es comisionado por la Universidad de Oxford en la Oficina del Gabinete del Reino Unido donde es director de evaluación de política en la Unidad de Estrategia. En Oxford el doctor Davies es director de ciencias sociales en el Departamento para la Educación Continua, y es miembro de la junta de gobierno del Kellogg College en Oxford. Dr. Philip Davies es graduado de las universidades de Londres, Oxford y California. Además, ha desarrollado investigaciones y cátedras en otras universidades del Reino Unido y los Estados Unidos. Fue responsable (con otros colegas de la Escuela de Medicina de la Universidad de Oxford) del desarrollo de un Programa de Maestría en Oxford sobre Cuidado de la Salud basado en la Evidencia. Dirigió ese programa durante los primeros años de operación. Recientemente, Philip Davies ha trabajado con colegas en la Gran Bretaña y Norteamérica para desarrollar políticas públicas basadas en la evidencia. Pertenece al comité de dirección de la Asociación Campbell, que prepara, mantiene y difunde las revisiones sistemáticas de los efectos de las intervenciones en educación, crimen y justicia y bienestar social. Es presidente del Grupo Educativo de la Asociación Campbell y presidente visitante honorario del Centro UK Cochrane. Entre sus publicaciones destacan: *Práctica en Salud Basada en la evidencia: Manual para Profesionistas en la Salud* (coautoría,

Edimburgo, Churchill Livingstone, 1999) y un ensayo en “¿Qué es lo que funciona? Evidencia para la Política Pública” (Bristol, Policy Press, 2000). En revistas recientes: “¿Qué es la Educación basada en la Evidencia?” (*British Journal of Educational Studies*, 47, 2, 108-120); “La relevancia de Revisiones Sistemáticas de la Política y Práctica Educativa” (*Oxford Review of Education*, 26, 3 y 4, 365 -378); y “La Asociación Campbell hace para la política pública lo que Cochrane hace para la Salud” (*British Medical Journal*, 323, 294-295).

Director of Policy Evaluation
The Strategy Unit
UK Cabinet Office
Room 4.37
Admiralty Arch
The Mall
Londres, SW1A 2WH
Reino Unido
Tel.: +44 207 276 1864
Fax: +44 207 276 1450
Correo electrónico: phil.davies@cabinet-office.x.gsi.gov.uk

* * *

Dr. Randall Eberts

Randall Eberts es director ejecutivo para la Investigación sobre el Empleo del W. E. Upjohn Institute. Sus investigaciones incluyen la evaluación de programas de empleo y capacitación. Actualmente el doctor Eberts incluye el desarrollo de modelos estadísticos que ayudan a identificar las necesidades de los demandantes de empleo para que puedan dirigirse más rápida y efectivamente a servicios que atiendan mejor sus necesidades. Él y sus colegas han desarrollado modelos similares en Michigan para generar un sistema de perfilación y servicios de recomendación y para un proyecto piloto de beneficiarios de la seguridad social y participantes en el programa, financiado por el Departamento para el Empleo en los Estados Unidos. Ha preparado también reportes para la Comisión Europea sobre la experiencia estadounidense con una primera identificación de las necesidades de los trabajadores y los servicios potenciales para el trabajo que permiten estimular el crecimiento económico en Europa. El también se ha asociado con el programa LEED de la OCDE para examinar el papel de las asociaciones en el desarrollo de la fuerza laboral y en el desarrollo económico. Ha publicado numerosos artículos en revistas académicas como *Review of Economics and Statistics*, *Journal of Labor Economics*, *Journal of Human Resources*, y *Economic Inquiry*. Es autor y editor de diversos libros. Fue profesor asociado de economía en la Universidad de Oregon, profesor visitante en la Universidad de Texas A&M,

asistente del vicepresidente y economista del Banco de la Reserva Federal de Cleveland, y jefe del equipo de economistas del Consejo de Asesores Económicos del Presidente. Randall Eberts recibió su doctorado por la Universidad de Northwestern.

W.E. Upjohn Institute for Employment Research
300 South Westnedge Avenue
Kalamazoo, MI 49007
USA
Tel.: +1 (616) 343-5541
Fax: +1 (616) 343-3308
Correo electrónico: EBERTS@we.upjohninstitute.org

* * *

Dra. Kris Hallberg

Kris Hallberg es economista líder en el Departamento de Evaluación de Operaciones del Banco Mundial, una unidad independiente dentro del Banco Mundial que le reporta directamente al Cuerpo de Directores Ejecutivos del Banco. Desde que se unió al banco en 1986, la doctora Hallberg se ha especializado en comercio y política industrial, desarrollo del sector privado y de pequeñas empresas, principalmente en América Latina, el Este de Europa, Asia Central y del Este. Actualmente es responsable de la evaluación de las operaciones de desarrollo del sector privado del Banco. Durante algunos años trabajó con el Comité de Agencias Donantes para el Desarrollo de las Pequeñas Empresas generando directrices para la intervención de los donadores en el desarrollo de servicios empresariales para las pequeñas empresas. Es miembro del grupo de trabajo de evaluación del impacto del Comité de Donantes. De 1991 a 1994, la doctora Hallberg fue la jefa de la oficina del Banco Mundial en Colombia. Antes de unirse al Banco Mundial estuvo en las facultades del Amherst College y Colby College. La doctora Hallberg posee un doctorado en economía por la Universidad de Winsconsin-Madison.

The World Bank
MSN H3-307
1818 H Street, N.W.
Washington, DC 20433
USA
Tel.: +1 (202) 458 5570
Fax: +1 (202) 522 3123
Correo electrónico: khallberg@worldbank.org

* * *

Profesor Edward W. (Ned) Hill

Edward W. (Ned) Hill es profesor y distinguido académico del Departamento Económico de Asuntos Urbanos en el Maxine Goodman Levin College de la Universidad de Cleveland. Hill es también consejero en jefe, no residente, del Centro de Política Urbana y Metropolitana de Brookings Institution. Editó *Economic Development Quarterly* de 1994 a 2004. Edward Hill fue premiado con el título de Profesor y Académico Distinguido en el otoño del 2001. Fue nombrado miembro del Consejo de Administración de la Sociedad del Zoológico de Cleveland en el 2000, designado al Consejo de Directores de la Westside Industrial Retention Network (WIRE-Net) en Cleveland, y la Sociedad de MEMS en Ohio. El gobernador Robert Taft, le designó las Tareas de Revitalización Urbana de la Fuerzas Militares en el otoño de 1999. Edward Hill y Harold Colman fueron premiados con el Premio Robertson de los editores de la revista *Urban Studies* en 1994. Es el autor de dos libros, coeditor de cuatro, autor de más de 60 artículos, ensayos y columnas. Su libro publicado en el 2001, *Ventaja Competitiva de Ohio: Manufacturando Productividad*, y otros artículos y comentarios, pueden consultarse en la página de internet: http://urban.csuohio.edu/faculty/ned_hill/site/index.htm.

College of Urban Affairs
Cleveland State University
Cleveland, OH 44115
USA
Tel.: +1 (216) 687 2174
Fax: +1 (216) 687 9277
Correo electrónico: Ned@urban.csuohio.edu

* * *

Dr. Peter Huber

Peter Huber es investigador del Instituto Austriaco de Investigación Económica con especialidad en análisis del mercado laboral regional. Anteriormente trabajó como investigador del Departamento de Economía del Instituto de Estudios Avanzados de Viena. Ha contribuido con diversos estudios para la Unión Europea. Sus estancias de investigación incluyen la Universidad Palacky en Olomouc, República Checa y Hochschule für Ökonomie en la Antigua República Democrática Alemana.

WIFO
Postfach 91
A-1103 Wien
Austria
Tel.: +43 1 798 2601 404
Fax: +43 1 798 9386
Correo electrónico: Peter.Huber@wifo.ac.at
Sitio web: www.wifo.ac.at/

Hon. Henry McLeish

Henry McLeish ocupó un cargo de elección popular durante 30 años en los niveles local, regional y nacional del gobierno. Fue primer ministro de Escocia y ex ministro para la Empresa y la Formación Continua en el recientemente creado Parlamento Escocés. Antes de eso fue ministro para el Traspaso de Competencias en el gobierno de Anthony Blair. Cuando ostentó ese cargo, fue responsable del programa piloto para la legislación del Parlamento Escocés a través de la Cámara de los Comunes y presidió el Comité que preparó el anteproyecto de programa de trabajo para el nuevo parlamento cuando inició sus funciones oficiales en julio de 1999. Esta experiencia se ha combinado con casi 13 años de trabajo en el gobierno local y en universidades tanto de Escocia como de los Estados Unidos.

El Sr. McLeish, graduado en planeación urbana, cuenta con una amplia trayectoria como político, profesionista y académico sobre los efectos económicos y sociales de las políticas basadas en áreas, en los niveles de gobierno local, regional y nacional. Sus intereses particulares son: el papel de la gobernabilidad en el desarrollo económico, el control efectivo y el seguimiento de la inversión, y cómo la política pública basada en la evidencia puede mejorarse con una evaluación efectiva. Henry McLeish tiene una considerable experiencia en los diferentes enfoques del desarrollo económico de Escocia, el Reino Unido y Europa. El también ha participado en visitas gubernamentales a China, Japón, Hong Kong y Estados Unidos para discutir las políticas económicas.

Sus actuales áreas de interés incluyen el papel de la gobernabilidad en el apalancamiento positivo del cambio económico; desarrollos territoriales y temas sobre ventajas económicas competitivas; el nuevo regionalismo en Europa y la contribución que el gobierno de transición puede hacer para mejorar la economía; y la contribución que el aprendizaje, la economía del conocimiento y la tecnología, pueden hacer al desarrollo económico.

49 George Street
Cellardyke, KY10 3AS
Reino Unido
Tel.: +44 1 333 313 080
Correo electrónico: h.b.mcleish.internet.com

* * *

Dr. Hugh Mosley

Hugh Mosley recibió su doctorado en ciencias políticas por la Universidad de Duke, Estados Unidos. Es investigador en jefe del Consejo sobre el Mercado Laboral y el Empleo de la unidad de investigaciones del Centro de Investigación

en Ciencias Sociales de Berlín (WZB), donde ha trabajado desde 1986. Ha hecho investigación comparativa e interdisciplinaria sobre una gran variedad de temas relacionados con la política del mercado laboral. Su trabajo más reciente trata temas sobre la implementación, especialmente reformas al servicio público de empleo, y la evaluación de políticas. Cuenta con una vasta obra publicada sobre su experiencia como consultor de la Comisión Europea y de otras organizaciones internacionales en temas del mercado laboral: *Effizienz der Arbeitämter* (Berlín, 2003) en colaboración con Holger Sültz y Günther Schmid; *Mercados Laborales, Género y Cambio Institucional* (Edward Elgar, 2000) co-editado con Jacqueline O'Reilly y Klaus Scömann; "¿Cómo pueden hacerse a las políticas activas más efectivas?", en coautoría con Jaap de Koning en Günther Sumid y Bernard Gazier, *Las dinámicas del pleno empleo. Integración Social en los Mercados Laborales de Transición* (Edward Elgar, 2002); *Política del Mercado Laboral y el Desempleo: Impacto y Procesos de Evaluación en Países Europeos Seleccionados* (Edward Elgar, 2001) co editado con Jaap de Koning.

Wissenschaftszentrum Berlin (WZB)
Reichpietschufer 50
D-10785 Berlin
Alemania
Tel.: +49 30 25491 112/126
Fax. +49 30 25491 100
Correo electrónico: mosley@wz.berlin.de

* * *

Profesora Alice Nakamura

Alice Nakamura es profesora en la Escuela de Negocios de la Universidad de Alberta. Dra. en economía por la Universidad Johns Hopkins, también posee un doctorado honorario en leyes por la Universidad de Western Ontario. Es copresidenta académica del CERF, el Foro de Investigación Canadiense para el Empleo, y ex presidenta de la Asociación Canadiense de Economía. Ha sido consejera en las áreas de políticas laborales, productividad, medición del desempeño y evaluación de los gobiernos en Canadá, los Estados Unidos y otras partes. Es miembro de diversos consejos de administración de estadísticas en Canadá. Sus principales áreas de investigación y sus publicaciones se ocupan de temas como la economía laboral, inmigración, econometría y medición de la productividad.

Es también presidenta del instituto que fundó www.CareerOwl.ca, un servicio de reclutamiento electrónico que comenzó gracias al apoyo de miembros de las facultades en las universidades canadienses que ayudaban a reducir los costos para que las empresas locales encontraran los talentos que necesitaban, y han

ayudado a su mantenimiento a través del apoyo fiscal de las universidades de Canadá.

Alberta School of Business
3-23 Business Building
University of Alberta
Edmonton, Alberta
T6G 2R6
Canadá
Tel.: +1 780 492 5824
+1 780 492 2457 – secretaria
Fax: +1 780 492 9924 or 492 3325
Correo electrónico: alice.nakamura@ualberta.ca

* * *

Sr. Alistair Nolan

Alistair Nolan ha trabajado con la OCDE desde julio de 1997, especializándose en aspectos de política pública orientada al área empresarial, con un enfoque en los vínculos entre la creación de empresas y el desarrollo de las economías locales y regionales.

El señor Nolan jugó un papel clave en la elaboración de la publicación emblemática de la OCDE de 1998 *Fostering Entrepreneurship* y también fue responsable de dos libros sobre incubación de negocios: *Incubación de Negocios: Estudios de Caso Internacionales* (1999) y *Buenas Prácticas en la Incubación de Negocios* (2000). Él también ha sido responsable de las recomendaciones de política de la OCDE sobre redes de negocios y clusters empresariales. Es autor del libro *Empresarialidad y Desarrollo Económico Local* (2003), el cual plantea detalladamente las directrices para los programas y políticas de los gobiernos centrales y locales. Hasta enero de 2004 Nolan fue también responsable del trabajo del programa LEED de la OCDE sobre métodos y práctica de la evaluación. Actualmente es co-administrador de las fases preparatorias de una posible evaluación cuantitativa de países miembros de la OCDE sobre habilidades en los adultos que busca dar luz sobre una extensa selección de preocupaciones de política macro y micro económicas. Antes de unirse a la OCDE, trabajó como miembro del pequeño grupo responsable de monitorear, evaluar y dar asistencia técnica al Programa de las Naciones Unidas para la Organización del Desarrollo Industrial. En este contexto fue responsable de evaluar proyectos y programas en campos que van desde la capacitación hasta la transferencia de tecnología, producción ambientalmente limpia y promoción de la inversión. Durante muchos años ocupó cargos en la investigación, política y el diseño de la cooperación técnica con UNIDO. Nolan

tiene una maestría por la Universidad de Cambridge en Economía y Política para el Desarrollo, también tiene un post grado en finanzas corporativas y en economía financiera. Ha realizado estudios en economía ambiental y en financiamiento a proyectos. Es candidato a obtener el doctorado en economía por la Universidad de Cambridge.

OCDE
2, rue André-Pascal
75775 Paris Cedex 16
Francia
Tel.: +33 1 45 24 1386
Correo electrónico: Alistair.Nolan@OECD.org

* * *

Dr. Eric S. Oldsman

Eric Oldsman es el fundador y presidente de Nexuss Associates, Inc. Ha dirigido numerosos proyectos para una gran variedad de clientes, incluyendo agencias federales y estatales, instituciones multilaterales, organizaciones sin fines de lucro y corporaciones privadas líderes en América Latina, Asia, Europa y los Estados Unidos. Muchos de los estudios se enfocan al diseño, gestión y evaluación de las iniciativas de desarrollo económico. En años recientes, ha evaluado programas como el Ben Franklin Technology Partnership (Estados Unidos), Centros de Desarrollo Empresarial (Argentina), Programa de Integración Industrial (México), Proyecto de Apoyo al Desarrollo de Mekong (Vietnam), NIST Asociación para la Ampliación Manufacturera (Estados Unidos), Instituto Robert C. Byrd en los Estados Unidos, y el Instituto Tailandés sobre Productividad (Tailandia).

Las evaluaciones se han enfocado en aspectos relacionados con el desarrollo organizacional, el rendimiento, la eficiencia operativa, autosuficiencia financiera y efectividad. Se han utilizado diversas técnicas en la estimación de los impactos, incluyendo estudios de caso basados en la teoría, encuestas de satisfacción al cliente y diseños cuasi experimentales. Antes de fundar Nexus Associates en 1991, el doctor Oldsman fue durante siete años un importante consultor con Arthur D. Little, Inc. En el equipo de PACT, Inc., fue responsable del proyecto para dar seguimiento y evaluar programas de desarrollo comunitario en África y América Latina. Cuenta con un doctorado en Política Pública por la Universidad de Harvard y es economista por la Universidad de Brown.

Nexus Associates, Inc.
68 Leonard Street
Belmont, MA 02478
USA

Tel.: +1 (617) 489 0311

Correo electrónico: oldsmann@nexus-associates.com

Página web: www.nexus-associates.com.

* * *

Dr. Christopher J. O'Leary

Christopher J. O'Leary es un importante economista en el Instituto para la Investigación sobre el Empleo W. E. Upjohn. Su investigación sobre el seguro de desempleo ha examinado las tasas de recontractación, los perfiles, la pertinencia de los subsidios, el nivel de experiencia. El ha evaluado programas de capacitación, subsidio a los salarios, trabajos públicos, autoempleo y servicios para el empleo para ministerios del trabajo en países en transición como Hungría, Polonia y China. Para el Departamento del Trabajo de los Estados Unidos trabaja con Randall Eberts para desarrollar un sistema de apoyo en la toma de decisiones de primera línea para los centros de apoyo a los profesionistas bajo el Decreto de Inversión en la Fuerza de Trabajo. Su investigación ha sido patrocinada por el Banco Mundial, la Oficina Internacional de Empleo, y por el centro de Desarrollo a los Recursos Humanos de Canadá. Sus artículos han aparecido en el *Journal of Human Resources*, *Journal of Policy Analysis and Management*, *International Labour Review*, *New England Economic Review*, *Economics of Transition*, y *Applied Economics*. Christopher O'Leary completó sus estudios universitarios en la Universidad de Massachusetts en Amherst y obtuvo un doctorado en economía por la Universidad de Arizona. En 1999 fue electo para la Academia Nacional de Seguridad Social.

W.E. Upjohn Institute for Employment Research

300 South Westnedge Avenue

Kalamazoo, MI 49007

USA

Tel.: +1 (616) 343-5541

Fax: +1 (616) 343-3308

Correo electrónico: OLEARY@we.upjohninst.org

* * *

Dr. Jonathan Potter

Jonathan Potter ha trabajado como economista en jefe del Programa para el Desarrollo Económico y el Empleo a nivel Local (LEED) de la OCDE desde 1997, y es responsable de las actividades de LEED sobre empresarialidad y evaluación. Ha sido autor y editor de diversos libros de la OCDE, incluyendo *Descentralización y Globalización. Implicaciones para los tomadores de Decisiones a Nivel Local*; *Flujos de Conocimiento Global y Desarrollo Económico y La Dimensión Local, del Estado de*

Bienestar al Trabajo. También coordina dos series de estudios, *Las reseñas sobre OCDE empresarialidad local* y la serie sobre *Inversión Extranjera Directa y Desarrollo Local*. El doctor Potter fue anteriormente consultor en jefe del Grupo Consultivo PA en el Reino Unido, se especializó en la evaluación de la política pública. Ha llevado a cabo numerosas evaluaciones para la Comisión Europea, gobierno central y agencias de desarrollo local y regional y ha dado asesoría para el desarrollo de metodologías de evaluación. El doctor Potter obtuvo su doctorado en economía por la Universidad de Cambridge.

OCDE
2, rue André-Pascal
75775 Paris Cedex 16
Francia
Tel.: +33 1 45 24 89 77
Fax: +33 1 45 24 16 68
Correo electrónico: Jonathan.Potter@OECD.org

* * *

Profesor Brian Robson

Brian Robson es Director del Centro para Estudios de Política Urbana (CUPS) de la Universidad de Manchester. Ha sido profesor de Geografía en Manchester desde 1977 y, previamente, en la Universidad de Cambridge. Como Director de CUPS ha emprendido numerosos contratos de investigación para departamentos del gobierno central, autoridades locales y otros patrocinadores, así como proyectos académicos para el Consejo de Investigación Económica y Social. Sus investigaciones se han enfocado principalmente en tres áreas: evaluación del impacto de las iniciativas de política urbana; desarrollo de indicadores de pobreza y exclusión social y orígenes de las inequidades regionales. Ha publicado ocho libros, importantes reportes y más de 100 artículos académicos. Ha tenido un importante papel en asesorar al gobierno en muchas de las iniciativas recientes de regeneración urbana: el desarrollo de un presupuesto único par la regeneración, corporaciones para el desarrollo urbano, compañías de regeneración urbana y, actualmente, renovación del Fondo de Mercado Inmobiliario. Fue miembro del grupo de trabajo de asesoría fiscal del Lord Rogers' Urban Task Force, y miembro del Consejo Gubernamental de Sondeo Urbano. En el 2000 fue premiado con la Medalla de Oro que otorga la Royal Geographical Society por su contribución a la política pública. Recientemente fue electo miembro honorario del Royal Town Planning Institute.

Department of Geography
Manchester University
Manchester, M13 9PL

Reino Unido
Tel.: +44 161 275 3639
Fax: +44 161 275 7878
Correo electrónico: Brian.Robson@man.ac.uk

* * *

Profesor Jeffrey A. Smith

Jeffrey Smith es profesor en Economía de la Universidad de Maryland. Recibió su doctorado en Economía por la Universidad de Chicago en 1996 y se unió a la facultad en 2001. Antes de mudarse a Maryland fue profesor asociado y catedrático en Capital Humano y Productividad en la Universidad de Ontario en Canadá. Recibió en 1997 el premio Polanyi de la provincia de Ontario que se da cada año a un joven y sobresaliente economista. Sus investigaciones se centran en métodos para la evaluación de programas sociales tales como los de capacitación para el trabajo de los más desfavorecidos. Ha escrito artículos sobre los efectos en el mercado laboral de la calidad universitaria y utiliza el tratamiento de reglas estadísticas para asignar las personas a los programas gubernamentales. Sus publicaciones recientes incluyen “Substitución y sesgo por la Pérdida de Información en los Experimentos Sociales: Estudio de un Experimento Social Influyente” (con James Heckman, Neil Omán y Michael Khoo), en *Quarterly Journal of Economics* 2000; “Economía y Econometría de los Programas para Activar el Mercado Laboral” (con James Hackman y Robert LaLonde) en *Handbook of Labor Economics*, vol. 3A 1999, y “La pérdida de Ganancias Pre-Programa y los Determinantes de la Participación en un Programa Social: Implicaciones para un Programa Simple de Evaluación de Estrategias (con James Heckman), en *Economic Journal* 1999 (Ganador del premio que otorga la Royal Economic Society por el mejor artículo publicado en la revista *Economic Journal* en 1999).

Department of Economics
University of Maryland
3105 Tydings Hall
College Park, MD 20742-7211
USA
Tel.: +1 (301) 405 3532
Fax: +1 (301) 405 3542
Correo electrónico: smith@econ.umd.edu
Página web: www.bsos.umd.edu/econ/faculty/Smith.htm.

* * *

Dr. George I. Treyz

George I. Treyz tiene un doctorado por la Universidad de Cornell y licenciatura por Princeton. Es profesor emérito de Economía por la Universidad de Massachussets en Amherst. Es autor del libro *Modelo de Simulación de la Economía Regional* y es autor y coautor de más de 25 artículos académicos publicados en revistas profesionales, incluyendo la *American Economic Review*, *Review of Economic Statistics*, e *International Regional Science Review*. En 1980, el Dr. Treyz fundó el Regional Economic Models, Inc. (REMI) para desarrollar e implementar modelos regionales macro y micro económicos para clientes públicos y privados. Estos modelos se utilizan para predecir y realizar análisis de política en la mayoría de las ciudades en los Estados Unidos, en gran parte de los gobiernos estatales, agencias del gobierno federal, universidades, y en agencias gubernamentales en áreas pequeñas y grandes. Un grupo de investigación dirigido actualmente por los doctores George y Frederick Treyz ha completado recientemente el último REMI Policy Insight® basado en parte en el prototipo REMI y publicado en noviembre de 2000 como tema del *Journal of Regional Science*. Está diseñado para utilizarse tanto en los Estados Unidos como en la Unión Europea.

President
Regional Economic Models, Inc.
306 Lincoln Avenue
Amherst, MA 01002
USA
Tel.: +1 (413) 549 1169
Fax: +1 (413) 549-1038
Correo electrónico: georgetreyz@remi.com
Página web: www.remi.com.

* * *

Profesor Robert Walker

Robert Walter es Profesor de Política Social por la Universidad de Nottingham y consejero de investigación en el Instituto de Estudios Fiscales. Es también consejero de la Real Sociedad de Artes y fue el director fundador del Centro para la Investigación de Política Social en la Universidad de Loughborough.

Es partidario de que la investigación de alta calidad debe utilizarse para informar y mejorar las políticas públicas en favor de todos. Realiza investigaciones relevantes sobre el desarrollo de políticas de bienestar en la Gran Bretaña y en otras sociedades industriales avanzadas. El también impulsa el diálogo con los diseñadores de políticas y con otros que quieren utilizar o apoyarse en la investigación para alcanzar un cambio positivo.

Ha dirigido investigaciones para diferentes instancias gubernamentales y para organismos internacionales desde 1983. Ha realizado importantes evaluaciones de diversas políticas en el Reino Unido, incluyendo el Nuevo Trato para la Personas con Discapacidad, el Seguro de Desempleo y el Fondo Social. Recientemente se ha involucrado en el meta análisis de los programas de bienestar para el trabajo en los Estados Unidos. Sus áreas de estudio y evaluación incluyen progresiones en el empleo y desempleo; la pobreza y la dinámica de la pobreza; la exclusión social; las aspiraciones de los niños, la dinámica familiar y las estrategias de presupuesto familiar. Ha publicado 17 libros y más de 40 informes de investigación.

Professor of Social Policy
 Department of Sociology and Social Policy
 University of Nottingham
 University Park
 Nottingham, NG7 2RD
 Reino Unido
 Tel.: +44 115 951 4546
 Fax: +44 115 951 5232
 Correo electrónico: Robert.Walker@nottingham.ac.uk

* * *

Dr. Stephen A. Wandner

Stephen Wandner ha sido Director de Investigaciones y experimentación del Departamento del Trabajo, Empleo y Administración de la Capacitación en los Estados Unidos desde enero de 1997. Ha iniciado y dirigido un gran número de investigaciones y proyectos experimentales relacionados con el Decreto de Inversión en la Mano de Obra: programas de capacitación, programas de intercambio, empleo público, programas para jóvenes y seguro de desempleo. Inició un proyecto que brinda capacitación a los microempresarios, asistencia técnica, transferencia de pagos y préstamos a una gran variedad de participantes en tres estados. Recientemente editó y publicó el libro *Focalizando los Servicios para el Empleo*.

En 1996 fue investigador visitante en jefe del Instituto Urbano en donde escribió y editó el libro *Seguro de Desempleo en los Estados Unidos: Análisis de Temas de Política*. Hasta 1996 fungió como director en funciones y subdirector de la Oficina de Legislación y Servicios de Actuaría del Servicio del Seguro de Desempleo.

También fue investigador titular en el Servicio del Seguro de Desempleo, donde coordinó ocho experimentos de gran escala que brindaron asistencia para la recolocación de receptores del seguro de desempleo. Dos de esos proyectos

brindaron las bases para la promulgación de la legislación federal sobre servicios para la recolocación de trabajadores y la asistencia para el auto empleo.

Obtuvo su doctorado en economía por la Universidad de Indiana.

US Department of Labor
200 Constitution Avenue, NW
Washington, DC 20210
USA
Tel.: +1 (202) 693 3663
Fax: +1 (202) 219 9074
Correo electrónico: WANDNER.Stephen@dol.gov

* * *

Andrea Westall

Andrea Westall es subdirectora de la Fundación Nueva Economía. Fue directora de la Unidad de Política en la Fundación para la Gestión Empresarial en la Escuela de Negocios de Londres. Anteriormente se desempeñó como investigadora consejera del Instituto para la Investigación en Política Pública. En el 2001 escribió un reporte sobre las empresas sociales y su contribución en políticas contra el desempleo, la salud, la exclusión financiera y la creación de mercados. Ha colaborado con las agencias de Desarrollo Regional de la Región Central de Inglaterra y del Noroeste, en la Unidad de Empresa Social (DTI) del Gobierno del Reino Unido y en la Coalición de la Empresa Social del Reino Unido. Actualmente evalúa los indicadores de impacto para las empresas sociales a través del proyecto EQUAL financiado por la Unión Europea. Andrea ha colaborado con la Unidad de Desarrollo e Innovación de la Oficina del Gabinete del Reino Unido donde revisa la reforma del sector de participación voluntaria. Asimismo, desarrolla una investigación sobre la generación de ingreso y la empresa social para la Fundación de Beneficencia y Apoyo y el Consejo Nacional de Organizaciones Voluntarias. Sus actividades abarcan evaluaciones sobre el papel de la empresa en la creación y apoyo de las actividades de regeneración, a través del Servicio de Apoyo a Pequeños Negocios del Departamento para el Comercio y la Industria en el Reino Unido.

New Economics Foundation
3 Jonathan Street
London, SE11 5NH
Reino Unido
Tel.: +44 207 820 6322
Fax: +44 207 820 6301
Correo electrónico: andrea.westall@neweconomics.org
Página web: www.neweconomics.org.

Dr. Robert A. Wilson

Robert Wilson es consejero en la Dirección de Investigación del Instituto para la Investigación sobre el Empleo en la Universidad de Warwick en el Reino Unido. Dicho Instituto es uno de los centros europeos líderes en investigación de campo del mercado laboral dedicado a las áreas de cambio ocupacional y desarrollo de habilidades. En particular, Robert Wilson encabeza el trabajo referente a los pronósticos sobre el mercado laboral y el comportamiento del mercado de trabajo; así como el cambio tecnológico y estructural y su impacto en el mercado laboral. Sus investigaciones incluyen estudios sectoriales que van desde la ingeniería hasta la construcción y los servicios de salud, analizando grupos específicos como científicos e ingenieros (aspectos sobre el papel que juegan dentro de las organizaciones, la educación, la capacitación y las situaciones laborales que les afectan en su conjunto). Asimismo, ha editado publicaciones relacionadas con pronósticos sobre empleos tales como *Pronósticos sobre el Empleo* (publicado por la Agencia de Desarrollo de Habilidades Sectoriales) *Proyecciones de Ocupaciones y Capacidades* (publicado por el Departamento de Educación y Habilidades) y la *Revisión de Economía y Empleo del Instituto*. Es autor de *Predicciones sobre el Empleo en la industria de la Construcción*; *El Servicio Nacional de Salud y el Mercado Laboral*; *Cambio Técnico: El papel de los Científicos e Ingenieros*; e *Investigación y Desarrollo de Estadísticas*. Ha sido miembro del Comité Permanente de Asesores Médicos y del Grupo de Investigación sobre Habilidades de la Fuerza de Trabajo.

Institute for Employment Research
 University of Warwick
 Coventry, CV4 7AL
 Reino Unido
 Tel.: +44 24 7652 4127
 Fax: +44 24 7652 4241
 Correo electrónico: r.a.wilson@warwick.ac.uk

* * *

Profesor Gary Woller

Gary Woller es profesor asociado de gestión pública en el Instituto de Administración Pública en Romney, en la Escuela Marrit de la Universidad Brigham Young donde imparte las asignaturas de Gestión Internacional del Desarrollo y Política Pública. Ha publicado numerosos artículos en revistas académicas y profesionales sobre desarrollo internacional y microfinanzas. Ha sido consultor de varias instituciones microfinancieras, bancos de desarrollo, organizaciones no gubernamentales de desarrollo internacional y firmas de consultoría sobre desarrollo en áreas de evaluación de impacto e investigación de

mercado. Actualmente es consultor en jefe de la Red de Promoción Educativa de Pequeñas Empresas, una asociación internacional de profesionistas con sede en Washington D.C., donde ha organizado cursos sobre evaluaciones de impacto y herramientas de investigación de mercados; coordinado proyectos de investigación bajo el auspicio de la Fundación Ford, y colaborado con docenas de instituciones de microfinanciamiento con propósitos de investigación y para desarrollar la capacidad institucional en tareas de evaluación de impactos e investigación de mercados. Gary Soller es cofundador y editor de la *Revista sobre Microfinanzas*. Obtuvo un doctorado por la Universidad de Rochester y la maestría y la licenciatura por la Universidad de Brigham Young.

Romney Institute of Public Management
766 TNRB
Brigham Young University
Provo, UT 84602
USA
Tel.: +1 (801) 378-4221
Correo electrónico: wollerg@yahoo.com

* * *

Dr. Ging Wong

Ging Wong es el director general de Administración, Regiones y Servicios de Correspondencia del Patrimonio Canadiense. Antes de eso fue director del área de Políticas de Capacidades dentro de la Iniciativa de Investigación de Política de la Oficina del Consejo Privado del Gobierno de Canadá, donde diseñó la investigación sobre la política, las capacidades funcionales y las redes en apoyo a las prioridades de política del gobierno de Canadá. Fue director de Evaluación Estratégica y Seguimiento del Desarrollo de Recursos Humanos en Canadá donde evaluó los programas sociales que absorben el 46 por ciento del presupuesto federal. Esto incluía evaluaciones a las pensiones públicas, estándares de desempeño, evaluación y seguimiento a las reformas del seguro para el empleo y los impactos en los individuos, comunidades y la economía, requeridos por los legisladores en el reporte anual al parlamento durante los años 1997-2001. Hizo una importante contribución a la práctica de la evaluación en Canadá al integrarla a la política estratégica durante su mandato, a través de una revisión innovadora del seguro de desempleo que mantuviera informadas las discusiones sobre las políticas públicas y ayudara a configurar las reformas legislativas. Anteriormente, inició diversos consejos sectoriales para el desarrollo de los recursos humanos y colaboró en la elaboración de la reforma a la Estrategia para el Desarrollo de la Fuerza Laboral en Canadá como jefe de la política para el empleo.

Ging Wong fue fundador del Foro Canadiense de Investigación sobre el Empleo (CERF), del que es miembro ejecutivo por la Asociación Canadiense de Economía. Formó parte de las principales redes de investigación: La Red Internacional Canadiense del Trabajo (CILN) y la Red de Investigación Occidental sobre Educación y Capacitación (WRNET). Ging Wong se ha mantenido como delegado canadiense del Comité Directivo del Programa LEED de la OCDE.

Antes de ser reclutado en el gobierno federal por el Departamento de Asuntos Exteriores, Ging Wong fue profesor en la Universidad de Calgary. Se tituló con honores en la Universidad de Calgary y realizó estudios de postgrado en la Universidad de Oxford, con disertaciones sobre economía laboral y relaciones industriales. Mantiene sus investigaciones como profesor adjunto en la Facultad de Negocios y Administración de la Universidad de Alberta y fue designado para un puesto similar en la Universidad Wuhan en China.

Canadian Heritage
25 Eddy Street, 11th Floor
Postal Locator: 25-11-0
Gatineau, Quebec
Canadá K1A 0M5
Tel.: +1 (819) 952 3514
Fax: +1 (819) 956 3645
Correo electrónico: ggwong@sympatico.ca

