Organisation for Economic Co-operation and Development

EDU/WKP(2020)18

Unclassified

English - Or. English

7 October 2020

DIRECTORATE FOR EDUCATION AND SKILLS

Assessing adults' skills on a global scale: A joint analysis of results from PIAAC and STEP

OECD Education Working Paper No. 230

François Keslair and Marco Paccagnella, OECD.

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

François Keslair (francois.keslair@oecd.org)
Marco Paccagnella (marco.paccagnella@oecd.org).

JT03466467

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

OECD EDUCATION WORKING PAPERS SERIES

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed herein are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome, and may be sent to the Directorate for Education and Skills, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Colombia was not an OECD Member at the time of preparation of this publication. Accordingly, Colombia does not appear in the list of OECD Members and is not included in the zone aggregates.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at http://www.oecd.org/termsandconditions.

Comment on the series is welcome, and should be sent to edu.contact@oecd.org.

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

---

www.oecd.org/edu/workingpapers

Copyright © OECD 2020

Abstract

This paper illustrates similarities and differences between two international surveys that assess adults' skills: the Programme for the International Assessment of Adult Competencies (PIAAC) and the Skills Towards Employment and Productivity (STEP) survey. In particular, the paper highlights the issues that can arise for researchers interested to jointly analyse the data from the two surveys and to compare their results. The paper finds that, in spite of the many similarities, important differences exist between PIAAC and STEP, both in the way the data are collected, and in the way the proficiency of respondents is estimated. These issues can indeed affect the cross-country comparability of results from the two surveys. There is instead little evidence that the literacy assessment used in the two surveys is not adequate to form a basis for a valid assessment of adults' proficiency on a global scale.

Résumé

Cet article illustre les similitudes et les différences entre deux enquêtes internationales qui évaluent les compétences des adultes : le Programme pour l'évaluation internationale des compétences des adultes (PIAAC) et l'Enquête sur les compétences pour l'emploi et la productivité (STEP). En particulier, le document met en évidence les problèmes qui peuvent se poser aux chercheurs souhaitant analyser conjointement les données des deux enquêtes et comparer leurs résultats. Le document démontre que, malgré les nombreuses similitudes, des différences importantes existent entre PIAAC et STEP, tant dans la façon dont les données sont collectées que dans la manière dont les compétences des répondants sont estimées. Ces différences peuvent en effet affecter la comparabilité des résultats des deux enquêtes entre les pays. Il y a cependant peu de preuves que l'évaluation de la littératie utilisée dans les deux enquêtes ne soit pas adéquate pour constituer une base pour une évaluation valide des compétences des adultes à l'échelle mondiale.

Table of contents

Tables

Figures

Boxes

1. Introduction

The Programme for the International Assessment of Adult Competencies (PIAAC) is a cross-national assessment of adult skills that was launched in 2008 by the OECD. Between 2008 and 2018, the PIAAC assessment has been administered in over 40 countries and economies over three rounds of data collection.

The Skills Towards Employability and Productivity (STEP) programme was designed by the World Bank to better understand the interplay between skills and employability in low- and middle-income countries. Developed in 2010-11, it has been implemented in 17 countries between 2012 and 2017.

The purpose of this paper is to illustrate the differences and the similarities between PIAAC and STEP, and to offer a joint analysis of the data they collect and the lessons that can be learnt from them.

PIAAC and STEP share many similarities. The main objective of both studies is to assess the skills of the adult population in order to better understand how skills are developed and maintained and how they contribute to adults' outcomes. To achieve this, they use very similar tools: a questionnaire administered in person by a trained interviewer to collect a wide range of socio-economic and demographic information from adults, and an assessment of the information-processing skills of respondents. Importantly, the assessments used in the two studies share many items and were designed to ensure that the results could be placed on the same scale.

A first important difference between PIAAC and STEP is that the former is primarily conceived as a cross-national comparative survey that could serve as an international benchmark for adults' proficiency in participating countries and economies. Ensuring the cross-country comparability of the results is at the core of the PIAAC methodology and study design. While STEP uses common instruments and survey procedures in the participating countries, the comparative focus is far less strong than in PIAAC. This is evidenced by the fact that the target populations differ between participating countries and that the results were mainly released in the form of country reports and country-specific databases (Roseth, Valerio and Gutiérrez, 2016[1]).

A second difference is that, as an initiative of the OECD, PIAAC is targeted primarily to OECD member countries and economies, including the world's most economically developed countries. STEP is administered by the World Bank and is therefore more addressed to the needs and specificities of low- and middle-income countries.

Box 1. Classifying countries by income level

The references to countries' income levels in this paper (e.g. as "high-income" or "low-income" countries) is based on the classification produced by the World Bank, which classifies countries in four income groups according to their Gross National Income [GNI; see (World Bank, 2019[2])]:

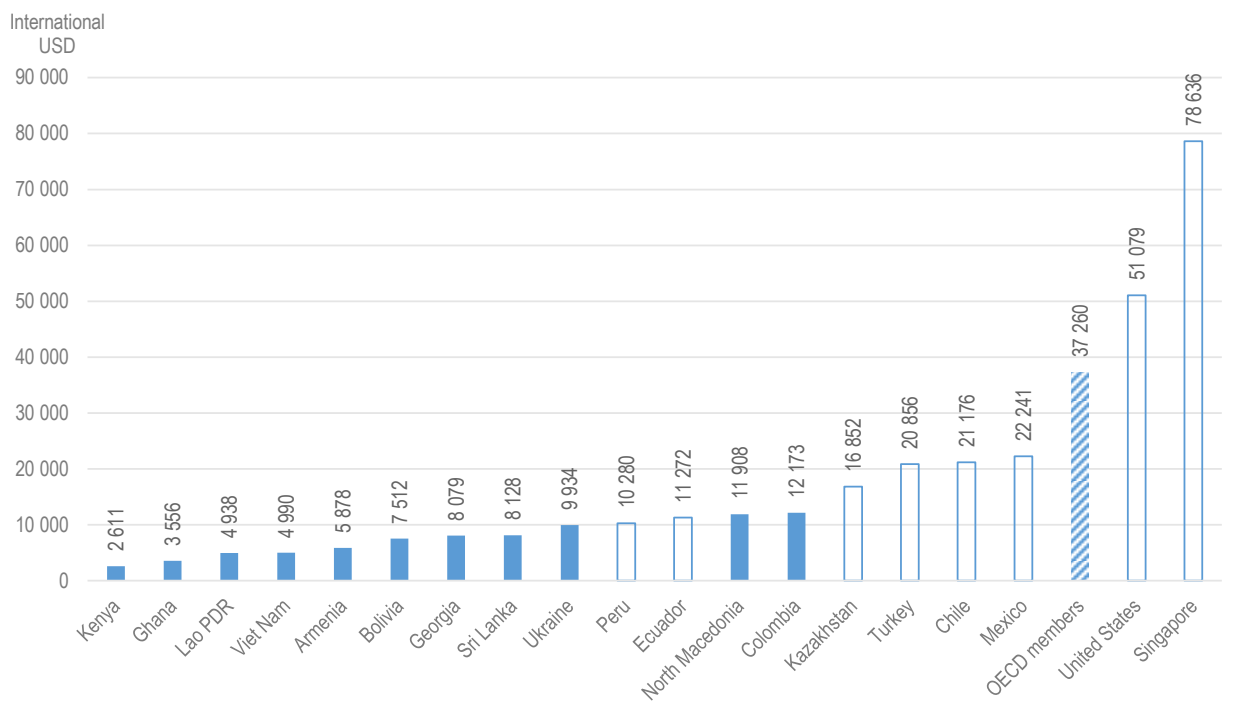
1. Low income, with a GNI below USD 1 026 (US Dollars)
2. Lower-middle income, with a GNI between USD 1 026 and USD 3 995
3. Upper-middle income, with a GNI between USD 3 996 and USD 12 375
4. High-income, with a GNI above USD 12 375.

All countries that participated in PIAAC are classified as high-income, with the exception of Ecuador, Kazakhstan, Mexico, Peru, the Russian Federation, and Turkey, which are all

upper-middle income. Indonesia is a lower-middle income countries, but the data were collected only in the capital city of Jakarta.

All countries that participated in STEP are either lower-middle income (Bolivia, Ghana, Kenya, Lao People's Democratic Republic [hereafter 'Lao PDR'], Ukraine and Viet Nam) or upper-middle income (Armenia, Colombia, Georgia, North Macedonia and Sri Lanka). Figure 1 illustrates this by plotting average GDP per capita over 2010-15 in some selected STEP and PIAAC countries.

Figure 1. Average levels of GDP per capita, 2010-2015

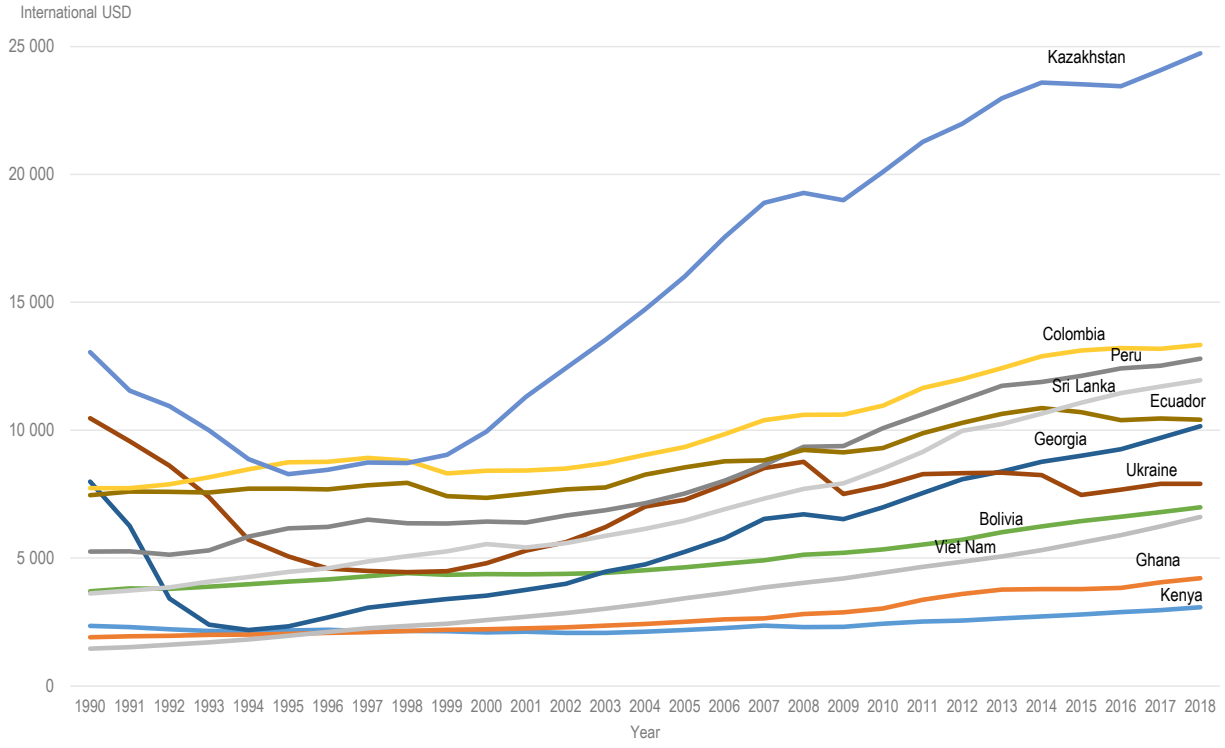


Notes: GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the US dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2011 international dollars.
Blue bars denote countries that participated in STEP and white bars denote countries that participated in PIAAC.
Countries are sorted by GDP per capita.
Source: World Bank ([3]) World Development Indicators. https://databank.worldbank.org/source/world-development-indicators (accessed on 28 October 2019).

Countries differ not only in terms of their current level of national income, but also in terms of the evolution of national income over time. This is relevant because the target populations of PIAAC and STEP are adults of working age. The proficiency of an adult is likely to be influenced more by their experiences of schooling and work over the entire lifecycle rather than by income level in the year of the survey. This is illustrated in Figure 2 and Figure 3, which show the evolution of GDP per capita and the educational attainment of adults aged 25 or more in 1970 and 2010 in a selected number of countries that

participated in either PIAAC or STEP[1]. The pathways and timing of phases of economic development in different countries are likely to influence the process of skills accumulation through education and labour market experiences. It is therefore important to take these structural differences between PIAAC and STEP countries into account when interpreting the data from the two studies.

Figure 2. Evolution of GDP per capita

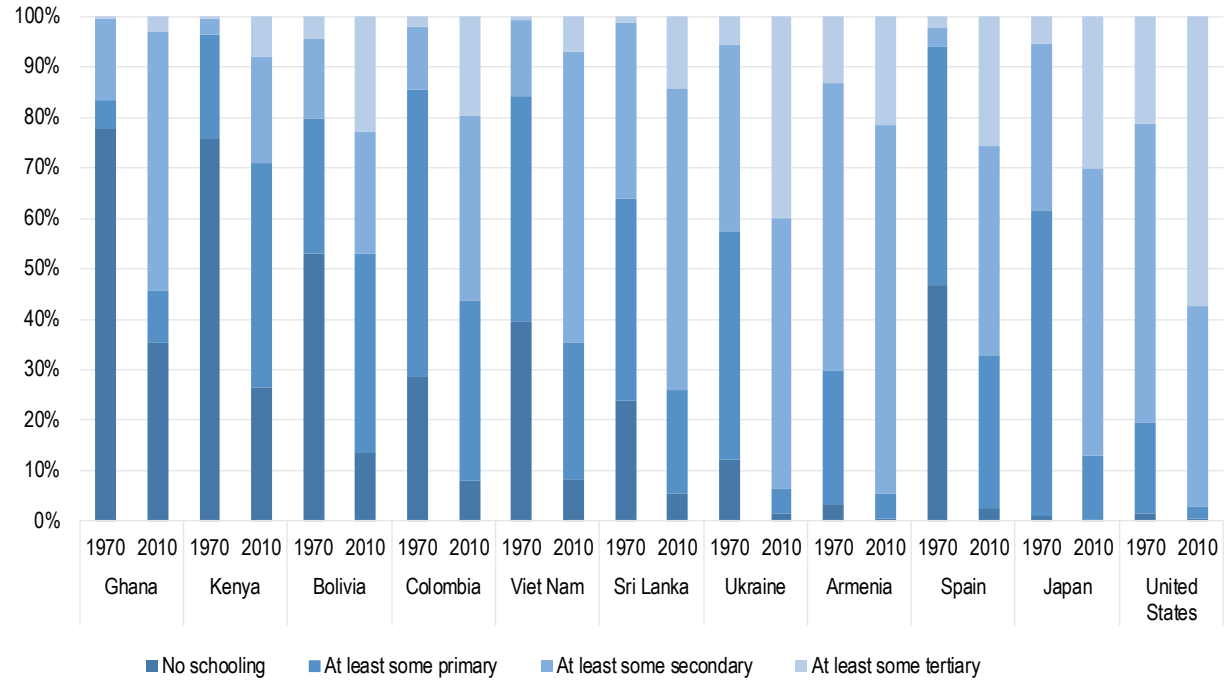


Note: GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the US dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2011 international dollars.
Source: World Bank ([3]) World Development Indicators, https://databank.worldbank.org/source/world-development-indicators (accessed on 28 October 2019).

[1] The choice of which countries and economies are represented in the graphs is partly driven by data availability.

Figure 3. Educational attainment in selected PIAAC and STEP countries

Adult population aged 25 years or more



Source: Barro and Lee (2013[4]), "A new data set of educational attainment in the world, 1950-2010", http://dx.doi.org/10.1016/j.jdeveco.2012.10.001.

In addition, PIAAC and STEP were designed to cater to the different characteristics and needs of their participating countries. Consequently, the adaptation to the local contexts generates differences in the instruments and in the surveys' administration procedures, which while minor in most cases, still have a bearing on the comparability of the final results.

The rest of the paper is structured as follows. Chapter 2 describes the content of the PIAAC and STEP surveys, highlighting differences in design, survey administration, and survey instruments. Chapter 3 discusses differences in scoring and scaling and how they may affect the estimates of skills proficiency. Chapter 4 presents a brief overview of estimated literacy proficiency in PIAAC and STEP participating countries and economies and analyses the relationship between skills and socio-demographic characteristics of participating adults, while Chapter 5 focuses on the relationship between skills and labour market outcomes. Chapter 6 concludes.

2. PIAAC and STEP: A comparison

This chapter describes the design of the PIAAC and STEP surveys and the instruments used for data collection, highlighting similarities and differences and pointing to how the differences affect the comparability of data collected in the two surveys. In order to collect comparable data, any cross-national survey should have a common design (including a similarly defined target population), and a survey protocol setting common rules and guidelines (Survey Research Center, 2016[5]). The countries and economies participating in PIAAC or STEP agreed to comply with a set of technical standards and guidelines,

specific to each programme (OECD, 2013[6]; Lyberg, Cibelli Hibben and Pennell, 2018[7]; World Bank, 2014[8]); coordination, supervision, and preparation of the survey instruments were centralised.

2.1. Survey and sampling design

2.1.1. Target population

While similar in many ways, there are important differences between the target populations for the two studies (see Table 1):

Table 1. Target population in PIAAC and STEP

	PIAAC	STEP
Age range	16-65 years	15-64 years
Residential status	Non-institutionalised usual residents of the geographic area covered	Non-institutionalised usual residents of the geographic area covered
Geographic coverage	The entire national territory. In a few cases, sub-national units	Urban areas. In some cases, the entire national territory
Exclusion	Up to 5% of the target population	Up to 5% of the target population

Note: In PIAAC, the target population is restricted to residents of Flanders in Belgium, of England and Northern Ireland in the United Kingdom and of Jakarta (Indonesia). In STEP, the target population includes people living in rural areas in Lao PDR and Sri Lanka.
Sources: OECD (2013[9]) *Technical Report of the Survey of Adult Skills (PIAAC)*, https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf and World Bank (2014[8]) "STEP skills measurement surveys: innovative tools for assessing skills", http://documents.worldbank.org/curated/en/516741468178736065/STEP-skills-measurement-surveys-innovative-tools-for-assessing-skills (accessed on 1 April 2020).

The most significant difference between the target population for PIAAC and STEP concerns its geographic coverage. The PIAAC target population consists of all non-institutionalised adults who reside in the country (usual place of residency is in the country) at the time of data collection. Only in three cases the target population was restricted to some regions or sub-national entities: in Belgium only adults living in Flanders were surveyed; in the case of the United Kingdom the survey covered only England and Northern Ireland; in Indonesia only the capital city Jakarta was covered. In contrast, almost all countries participating in STEP restricted the target population to residents of 'urban areas' and only Lao PDR and Sri Lanka surveyed adults living in rural areas.

These differences in the geographic coverage are one of the most important caveats to be kept in mind when comparing results both across STEP countries and between PIAAC and STEP.

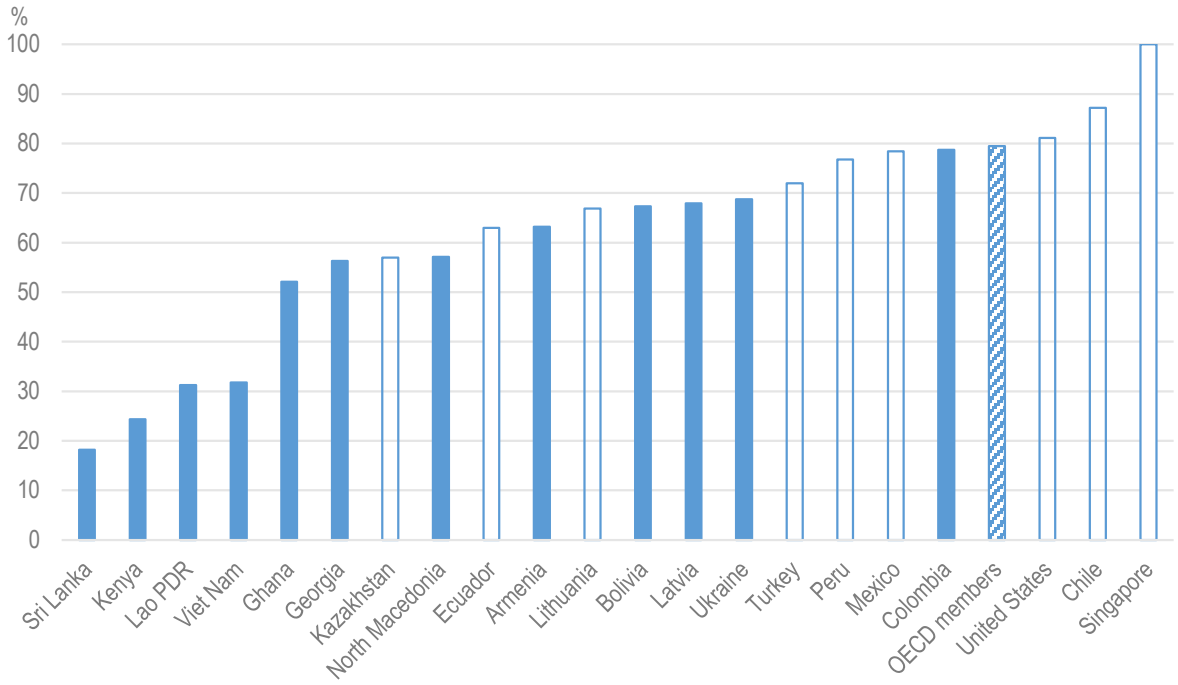
The extent to which restricting the geographic coverage to urban areas affects the results of the survey depends on how different adults living in urban and rural areas are, especially in terms of characteristics that are related to the cognitive skills assessed in STEP. Unfortunately, the scarcity of official statistics at the local or regional level, particularly in the countries participating in STEP, makes it very difficult to assess the extent to which the population in urban areas differs from that of the rural regions and, therefore, to estimate the extent to which these differences are likely to influence the level and distribution of adults' information-processing skills.

Urban-rural differences in terms of economic and social outcomes such as wages, consumption patterns, or educational attainment tend to be larger in low and middle-income countries (Puga, 2010[10]; Young, 2013[11]): adults living in rural areas are often less

educated than those living in urban areas (Figure 5), and are more frequently employed in sectors (like agriculture), requiring lower levels of the information-processing skills assessed in PIAAC and STEP.

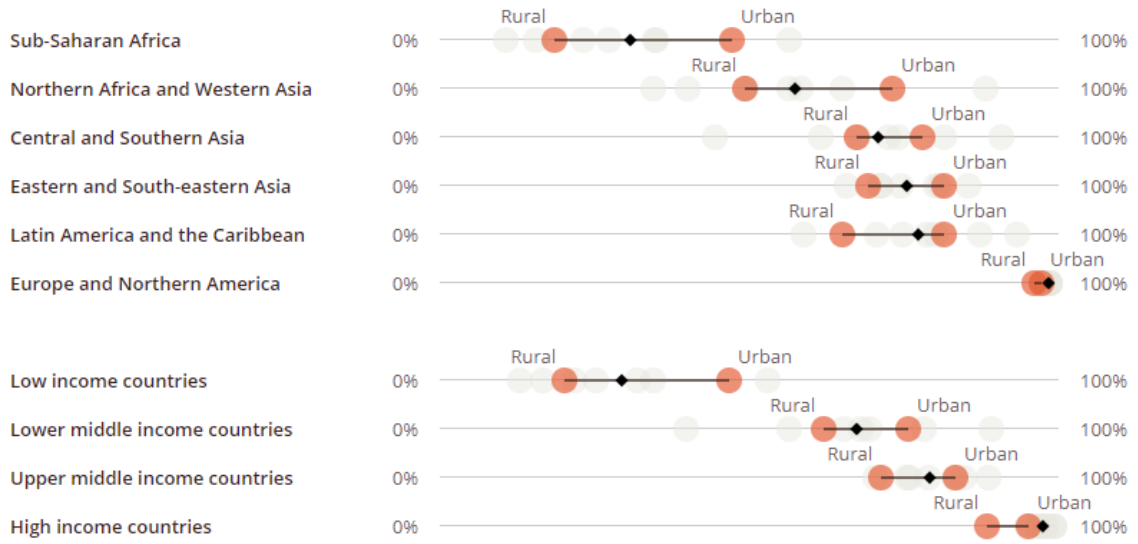
In low and middle-income countries a larger share of the population tend to live in rural areas compared to high-income countries (Figure 4). Indeed, the process of economic development is historically characterised by an increase in the share of the population migrating from rural to urban areas. Cities are normally recognised as a driver of economic development, and are often among the most productive areas of a country.

Figure 4. Share of population living in urban areas



Notes: Urban population refers to people living in urban areas as defined by national statistical offices. The data are collected and smoothed by United Nations Population Division. The data refer to 2012. Blue bars refer to countries that participated in STEP, white bars refer to countries that participated in PIAAC.
Source: World Bank ([3]) World Development Indicators. https://databank.worldbank.org/source/world-development-indicators (accessed on 28 October 2019).

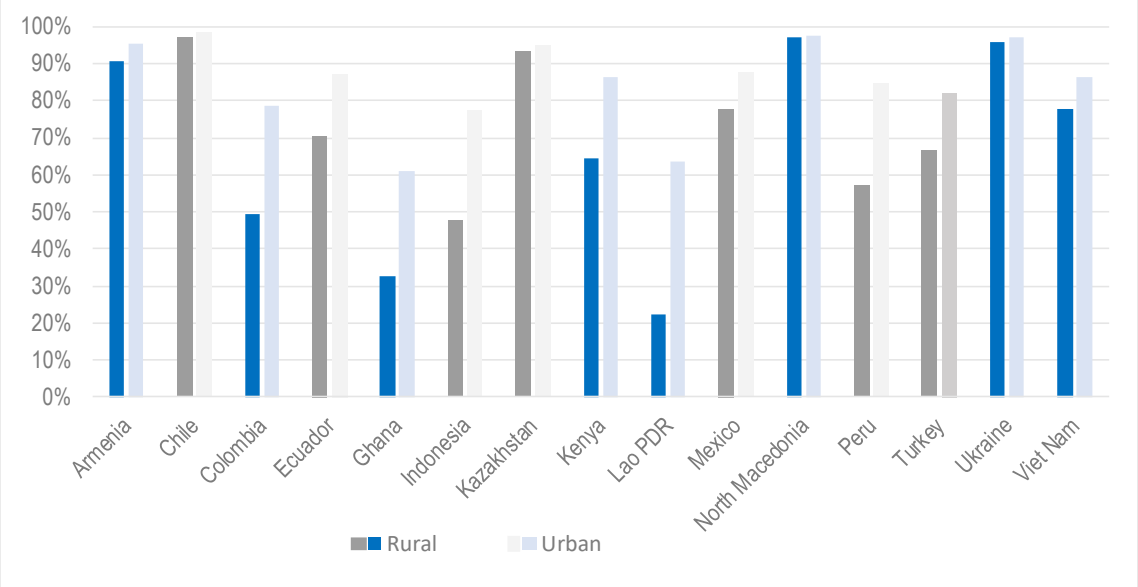
Figure 5. Rural-urban gaps in rates of completion of lower-secondary education



Note: People aged 15-24 years, latest available data.
Source: UNESCO ([12]) World Inequality Database on Education, https://unesco-wide-production.herokuapp.com/ (accessed on 1 April 2020).

A more detailed picture of rates of completion of lower-secondary education for young adults in countries that participated in PIAAC and STEP is given in Figure 6. Among the countries that took part in PIAAC, the figure reports those with the lowest levels of GPD per capita, in order to make the figure easier to read and, more importantly, to compare countries at similar stages of economic development.

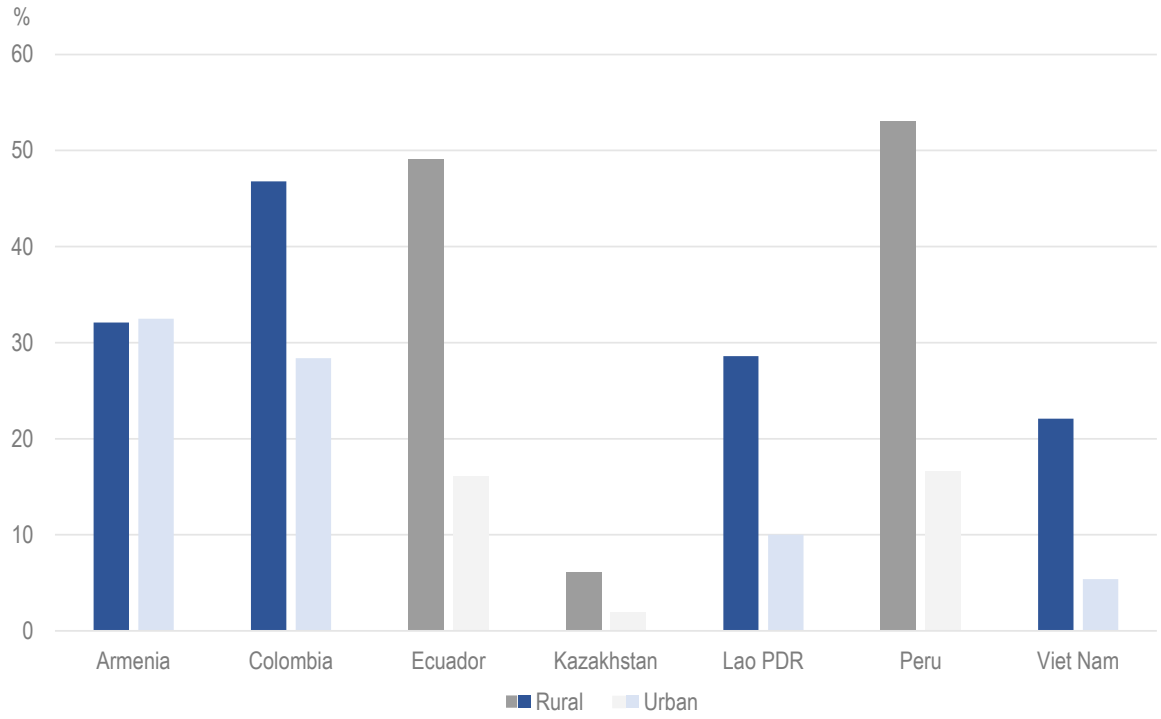
Figure 6. Completion of lower-secondary education in PIAAC and STEP countries



Notes: People aged 15-24 years, latest available data. Countries that participated in STEP are represented in shades of blue, countries that participated in PIAAC are represented in grey and white.
Source: UNESCO ([12]) World Inequality Database on Education, https://unesco-wide-production.herokuapp.com/ (accessed on 1 April 2020).

In many countries, the incidence of poverty also varies greatly between urban and rural regions. Figure 7 displays urban and rural poverty rates in a few selected PIAAC and STEP countries for which recent data are available.

Figure 7. Poverty rates in PIAAC and STEP countries, by location



Note: Urban and Rural Poverty Headcount Ratios at National Poverty lines (% of urban and rural population) in PIAAC and STEP countries and economies for which data are available.Data refers to 2012. Countries that participated in STEP are represented in shades of blue, countries and economies that participated in PIAAC are represented in grey and white.
Source: World Bank ([3]) World Development Indicators. https://databank.worldbank.org/source/world-development-indicators (accessed on 28 October 2019).

Finally, it should be noted that the definition of "urban areas" used in STEP was country-specific, decided by national statistical offices. This means that, even within STEP, some caution needs to be taken in directly comparing results across participating countries. In Bolivia, for example, the survey was restricted to four cities (La Paz, El Alto, Cochabamba, Santa Cruz de la Sierra); in Colombia 13 major metropolitan areas were covered (Bogota, Medellin, Cali, Baranquilla, Bucaramanga, Cucuta, Cartagena, Pasto, Ibague, Pereira, Manizales, Monteira, and Villavicencio); in Viet Nam, coverage was restricted to urban areas of Ha Noi and Ho Chi Minh City.

2.1.2. Sampling design

In both PIAAC and STEP, technical standards and guidelines define the broad features of the sampling designs that participating countries are expected to follow as well as the procedures for selecting individual respondents. In both studies, countries are required to implement a probability-based sample in which each member of the target population has a known (and non-zero) probability of selection. Each participating country usually has a unique sample design. In both PIAAC and STEP, participating countries and economies

usually implemented a complex design based on multi-stage sampling. In particular, households are selected first, and then an eligible adult is randomly selected within the household. A few countries and economies that participated in PIAAC, however, implemented a one-stage sampling design (these was the case in Austria, Flanders (Belgium), Denmark, Estonia, Finland, Netherlands, Norway, Singapore, Sweden, and of large localities in Israel). The PIAAC standard sample design is a self-weighting design, where each sample person (or household) has an equal probability of selection. In STEP, data weighting was carried out by STEP survey methodologists to ensure consistency across sampling strategies.

In both surveys, the technical standards and guidelines specified a number of criteria to check the quality of the collected data and the adherence to the sampling guidelines. A common indicator of data quality is response rates: low response rates can be an indication that the data collected are not representative of the target population, but rather of the selected sub-population of adults that agreed to participate in the survey. To mitigate concerns about selection bias, PIAAC countries and economies were required to perform a nonresponse bias analysis in case response rates were below 70%. Table 2 shows response rates in PIAAC and STEP participating countries and economies. The variation in response rates was greater in STEP than in PIAAC, largely due to several countries having very high response rates (over 90%). These outliers aside, response rates were broadly similar in the two studies.

Table 3 shows instead how much the STEP sample is actually representative of the national population, by comparing the size of the STEP target population with the number of people aged 15-64 living in the country. These coverage rates vary substantially, from about 10% in Viet Nam to 98% in Sri Lanka, and in most cases are below 60%. In PIAAC, the target population coincides with the national population, so that, by design, at least a 95% coverage of the national population is guaranteed for countries and economies that participated in PIAAC.

Table 2. Response rates in PIAAC and STEP

Countries and economies	Year	Reading Component	Coverage	Response Rate
PIAAC				
Australia	2011/12	Yes	National territory	71%
Austria	2011/12	Yes	National territory	53%
Canada[1]	2011/12	Yes	National territory	58%
Chile	2014/15	Yes	National territory	66%
Cyprus[2]	2011/12	Yes	National territory	73%
Czech Republic	2011/12	Yes	National territory	66%
Denmark	2011/12	Yes	National territory	50%
Ecuador	2017	Yes	National territory	69%
England (United Kingdom)	2011/12	Yes	Sub-national entity	59%
Estonia	2011/12	Yes	National territory	63%
Finland	2011/12	No	National territory	66%
Flanders (Belgium)	2011/12	Yes	Sub-national entity	62%
France	2011/12	No	National territory	67%
Germany	2011/12	Yes	National territory	55%
Greece	2014/15	No	National territory	51%
Hungary	2017	Yes	National territory	58%
Ireland	2011/12	Yes	National territory	72%
Israel	2014/15	Yes	National territory	61%
Italy	2011/12	Yes	National territory	56%

Japan	2011/12	No	National territory	50%
Jakarta (Indonesia)	2014/15	Yes	Sub-national entity	80%
Kazakhstan	2017	Yes	National territory	74%
Korea	2011/12	Yes	National territory	75%
Lithuania	2014/15	Yes	National territory	54%
Mexico	2017	Yes	National territory	71%
Netherlands	2011/12	Yes	National territory	51%
New Zealand	2014/15	Yes	National territory	63%
Northern Ireland (United Kingdom)	2011/12	Yes	Sub-national entity	65%
Norway	2011/12	Yes	National territory	62%
Peru	2017	Yes	National territory	83%
Poland	2011/12	Yes	National territory	54%
Russian Federation[3]	2011/12	No	National territory	52%
Singapore	2014/15	Yes	National territory	63%
Slovak Republic	2011/12	Yes	National territory	66%
Slovenia	2014/15	Yes	National territory	62%
Spain	2011/12	Yes	National territory	48%
Sweden	2011/12	Yes	National territory	45%
Turkey	2014/15	Yes	National territory	81%
United States	2011/12 and 2017	Yes	National territory	70%
STEP				
Armenia	2013	Yes	Urban areas only	50%
Bolivia	2012	Yes	Urban areas only	43%
Colombia	2012	Yes	Urban areas only	48%
Georgia	2013	Yes	Urban areas only	63%
Ghana	2013	Yes	Urban areas only	83%
Kenya	2013	Yes	Urban areas only	92%
Lao PDR	2012	Yes	National territory	95%
North Macedonia	2013	Yes	Urban areas only	67%
Philippines	2015	Yes	Urban areas only	95%
Sri Lanka	2012	Yes	National territory	63%
Ukraine	2012	Yes	Urban areas only	61%
Viet Nam	2012	Yes	Urban areas only	62%
Yunnan Province (People's Republic of China)	2012	Yes	Urban areas only	98%

Notes:
1. To account for multiple sampling frames and to provide an indication of nonresponse bias, nonresponse to the parent samples were reflected in Canada's PIAAC overall response rate computation. It was decided that individual response rates at the screener, Background Questionnaire and assessment stages would not be reported.
2. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue."
Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
3. Users should note that the sample for the Russian Federation does not include the population of the Moscow municipal region. The data published, therefore, do not represent the entire resident population aged 16-65 years in Russia but rather the population of Russia excluding the population residing in the Moscow municipal area.
Sources: Gaëlle et al. (2014[13]), "STEP Skills Measurement Surveys. Innovative Tools for Assessing Skills" and OECD (2019[14]), *The Survey of Adult Skills: Reader's Companion, Third Edition*, https://dx.doi.org/10.1787/f70238c7-en.

Table 3. Coverage of the national population in STEP countries

	Population covered in STEP	Population aged 15-64 in the country	Coverage of the national population by STEP
Armenia	1 185 359	2 014 458	58.8%
Bolivia	2 189 359	6 222 544	35.2%
Colombia	12 795 606	30 848 435	41.5%
Georgia	1 346 787	2 503 452	53.8%
Ghana	5 822 780	15 538 692	37.5%
Kenya	6 381 235	25 300 601	25.2%
Lao PDR	2 962 017	3 938 143	75.2%
North Macedonia	689 287	1 470 392	46.9%
Sri Lanka	13 424 281	13 618 903	98.6%
Ukraine	23 791 492	31 831 454	74.7%
Viet Nam	6 120 368	63 212 277	9.7%

Notes: The population covered by STEP is computed as the sum of final survey weights in the STEP database. The population aged 15-64 in the country is taken from the Health, Nutrition and Population Statistics database of the World Bank and refers to the year in which STEP was administered.
Source: World Bank ([15]) STEP Skills Measurement Program https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020) and World Bank ([16]) Health, Nutrition and Population Statistics Database, https://databank.worldbank.org/source/health-nutrition-and-population-statistics (accessed on 1 April 2020).

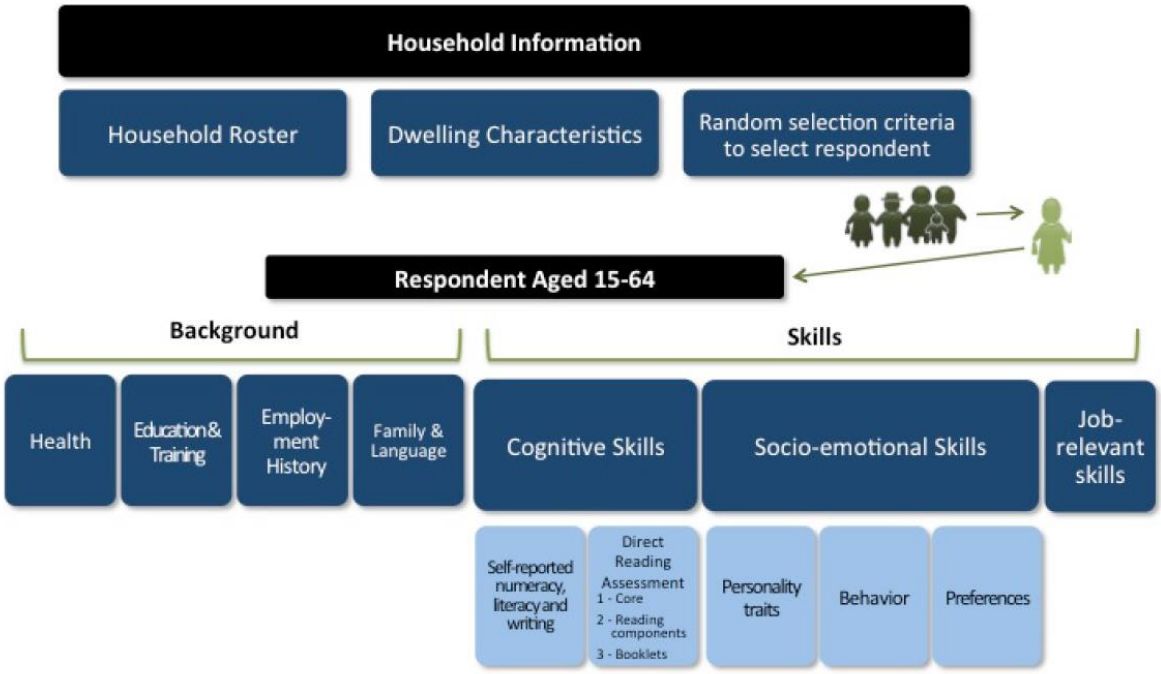
2.2. Survey instruments and assessment design

PIAAC and STEP adopted broadly similar approaches but slightly different survey designs. Data collection in both studies was conducted through personal interview: participating adults were visited in their household by trained interviewers and were administered a Background Questionnaire (BQ) and a cognitive assessment. While sharing much in common, there are considerable differences in the design of the two assessments. In particular, PIAAC was designed primarily as a computer-based assessment and tested three skills domains (literacy, numeracy and problem solving in technology-rich environments). Participating adults were allowed to opt for a paper-based assessment of literacy and numeracy, whose results were established to be equivalent to those of the computer-based assessment.

STEP, in contrast, assessed only one domain (literacy) using the same paper-based assessment instruments adopted in PIAAC (only in Colombia and Kenya the background interviews were carried out with the assistance of a computer). Moreover, not all countries that participated in STEP administered the full literacy assessment, which means that information on the literacy proficiency of the sampled population comparable to that available for PIAAC countries and economies is available only for Armenia, Bolivia, Colombia, Georgia, Ghana, Kenya, Ukraine and Viet Nam. The assessment designs of the two studies are summarised in Figure 8 and Figure 9.

In STEP the Background Questionnaire consisted of a first section collecting household-level information, and a second part collecting individual-level information from a randomly selected household member. The selection of the individual respondent within eligible household member was done by the interviewer. After completing the Background Questionnaire, the selected person would also be administered the cognitive assessment, which was essentially modelled on the paper-based version of the PIAAC literacy assessment: the two assessments shared many items and were scaled using a similar methodology, so that their results could be compared on a same scale.

Figure 8. STEP survey design

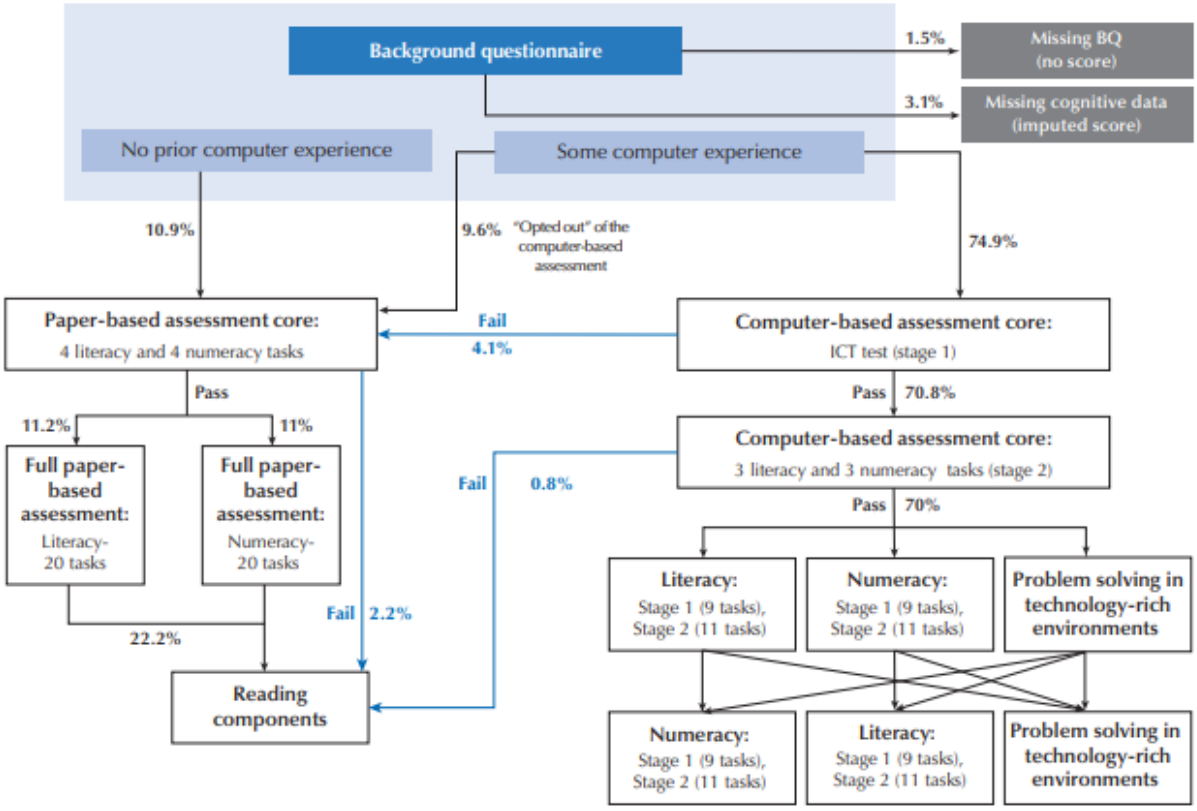


Source: World Bank (2014[8]), "STEP skills measurement surveys: innovative tools for assessing skills", http://documents.worldbank.org/curated/en/516741468178736065/STEP-skills-measurement-surveys-innovative-tools-for-assessing-skills (accessed on 1 April 2020).

The PIAAC BQ (and assessment) was only addressed to the randomly selected household member (although some information on other members of the household was also collected). The PIAAC assessment design was more complex, as it aimed to assess proficiency in three domains: literacy, numeracy, and problem solving in technology-rich environments (PSTRE). The assessment was designed to be primarily computer-based, but adults without prior computer experience, or who failed to show sufficient ICT proficiency, or that simply refused to take the assessment on a computer, were administered a paper-based assessment covering literacy and numeracy only.

In both PIAAC and STEP, a so-called "core" assessment was used as a screener: adults who did not pass this test were not administered the normal assessment, yet they received a score, estimated on the basis of their background characteristics. A separate assessment of reading components, designed to test some prerequisites of literacy (such as word comprehension and reading fluency) was administered to all STEP participants; in PIAAC, it was only administered to participants that, for any reason, undertook the paper-based branch of the assessment. A more detailed description of the two assessments is contained in the next section. The issues surrounding the scores of respondents who fail the core assessment are discussed in Chapter 3.

Figure 9. PIAAC assessment design



Note: The percentages presented in this diagram are based on the average of the OECD countries and economies participating in the Survey of Adult Skills (PIAAC).
Source: OECD (2019[14]), *The Survey of Adult Skills: Reader's Companion, Third Edition*, https://dx.doi.org/10.1787/f70238c7-en.

2.2.1. Background Questionnaire

Background questionnaires (BQ) are important components of large-scale surveys like PIAAC and STEP: their purpose is to collect information regarding aspects of respondents' demographic characteristics, background, and current education participation, work and living arrangements in order to better understand how and in which context people acquire the skills they demonstrate in the assessment. Without such contextual information, the results of the assessment would be much less useful, and much less relevant from both an academic and a policy perspective.

The PIAAC and STEP BQs have similar structures, but there are also many differences in the exact questions that are asked (even on the same topics). These differences limit the direct comparability and the joint analysis of the data. Both BQs include modules containing questions on the following broad topics:

- Characteristics of the household
- Education and training
- Health
- Employment
- Self-reported use of skills at home and on the workplace
- Personality, behaviours and preferences
- Language and family background.

Across the two BQs, the greatest commonality is seen in the modules covering education and training, employment, and language and family background. In both surveys, the Education module contains information on the highest qualification attained (classified according to the ISCED97 international classification). Comparable information is also available on field of study and on incomplete education. The Employment module contains questions on the type of job, occupation (coded accorded to the ISCO classification), job-searching strategies (for the unemployed), hours worked and wages. In the Language and Family Background Module, parental education can be used in both surveys as a proxy for socio-economic background.

The greatest differences are observed in the area of Health, Personality, Behaviours and Preferences. In these domains, there is little in common between the two studies. As far as Health is concerned, PIAAC only asks participants to provide a self-assessment of their overall health status. In contrast, STEP has a more extended set of questions covering overall life satisfaction, height, weight and the presence and severity of chronic health problems. The STEP module on Personality, Behaviour and Preferences has also a large set of items covering personality traits (a version of the Big Five personality inventory) and attitudes towards risk. The PIAAC BQ includes a much more limited set of items related to this domain with items on attitudes toward learning, trust, volunteering, and political efficacy.

Finally, both surveys collect information on the use of a wide range of skills at work and in everyday life. This includes both cognitive skills (related to the skills measured through the direct assessment) and other job-relevant skills, such as ICT skills, interpersonal skills, or problem solving skills. While the questions used to elicit this information can appear similar at first sight, there are important differences in the approaches used that mean that the information collected is not fully comparable. In PIAAC, information on skills use is elicited by asking respondents the frequency at which they perform certain tasks. For instance, to assess the use of reading skills, respondents are asked how often they engage in various forms of reading: reading books, reading manuals, reading emails, reading newspaper articles, and so on. Similarly, in terms of the use of interpersonal skills, respondents are asked how often their job requires them to persuade, influence, or negotiate with people. Borrowing items from the survey of Skills, Technology and Management Practice (STAMP; see Handel (2016[17]; 2008[18])), STEP attempts to measure not only the frequency, but also the complexity of the various tasks people carry out at work. As a consequence, many questions, while being similar to those asked in PIAAC, are often framed in a slightly different way, and have different response categories, making a direct comparison quite challenging. Table 4 exemplifies this by comparing how PIAAC and STEP elicit the use of reading skills.

The use of reading skills in the workplace is covered in similar ways in the two surveys, as participants report on their engagement with a similar set of reading tasks. However, there are differences in the approach to their measurement. The PIAAC BQ asks only about the frequency of performing these tasks. In contrast, in STEP, respondents are asked whether they undertake these tasks and then are asked about the complexity of these tasks defined in terms of the length of the material they normally read.

Table 4. Use of reading skills in the PIAAC and STEP questionnaires

PIAAC			STEP		
Question stem	Items	Response category	Question stem	Items	Response category
In your job, how often do you read…	letters, memos, emails?	1. Never; 2.Less than once a month; 3. Less than once a week; 4. At least once a week; 5. Everyday		Do you read anything at work?	Yes/No
	articles in newspapers or magazines?		As a regular part of your job, do you have to read…	forms?	Yes/No
	articles in professional journals?			bills or financial statements?	
	books?			newspapers or magazines?	
	manuals or reference materials?			instruction manuals?	
	bills, invoices, bank statements?			books?	
	diagrams, maps, schematics?			reports?	
				other?	
				Among the things that you normally read at this work, what is the size of the longest document that you read?	1. 1 page or less; 2.Between 2 and 5 pages; 3. Between 6 and 10 pages; 4. Between 11 and 25 pages; 5. More than 25 pages

Similarly, when assessing engagement with numeracy, PIAAC asks about the frequency of performing certain tasks. STEP instead only asks whether or not respondents perform certain tasks (binary response) and then classifies (subjectively) the tasks by levels of complexity.

Table 5 compares the way the two surveys elicit the use of interpersonal skills. PIAAC again asks about the frequency of engaging with different tasks or practices. The different items can then be combined to compute a latent Index of use of influencing skills at work. The STEP BQ contains two items relating to the use of interpersonal skills at work. The first asks about the frequency of interactions with colleagues or co-workers, using the same response categories as PIAAC. The second asks whether the respondent has to interact with people outside the firm (e.g. clients); in case the answer is positive, the respondent has to rate the intensity of these interactions on an involvement scale ranging from 1 to 10.

Table 5. Use of interpersonal skills in the PIAAC and STEP questionnaires

PIAAC			STEP	
Question stem	Items	Response category	Items	Response category
How often does your job usually involve…	instructing, training or teaching people, individually or in groups?	1. Never; 2.Less than once a month; 3. Less than once a week; 4. At least once a week; 5. Everyday	As part of this work, how frequently do you spend your time co-operating or collaborating with co-workers?	1. Never; 2.Less than once a month; 3. Less than once a week; 4. At least once a week; 5. Everyday
	making speeches or giving presentations in front of five or more people?		As part of this work, how frequently do you spend your time co-operating or collaborating with co-workers?	Yes/No. If Yes, rate the involvement on a scale from 1 to 10
	advising people?			
	planning the activities of others?			
	persuading or influencing people?			
	negotiating with people either inside or outside your firm or organisation?			

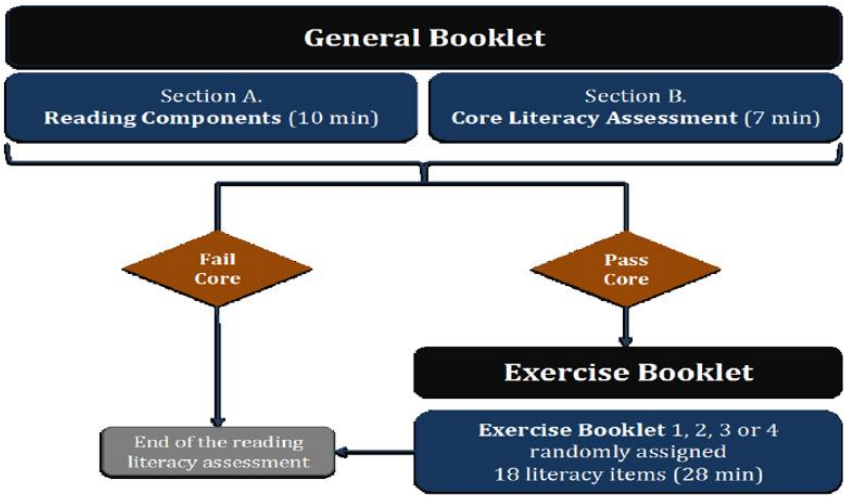
Note: The few examples reported above show that, while the two surveys aim to assess or elicit very similar concepts, the use of different instruments complicates any joint or comparative analysis, which would require some sort of harmonisation to make the information directly comparable.

2.2.2. Cognitive assessment

PIAAC and STEP assess the cognitive skills of participating adults using closely related instruments. In fact, STEP uses an enhanced version of the paper-based literacy assessment used in PIAAC. The design of the STEP assessment is depicted in Figure 10. All respondents first undertake an assessment of reading components (designed to assess basic reading fluency and passage comprehension) and a core literacy assessment, which serves as a kind of screener: only respondents who pass the core are administered the full reading literacy assessment, composed of 28 items distributed in different booklets.

PIAAC uses a similar but much more complex design (Figure 9). First of all, PIAAC screens respondents into a computer- or paper-based assessment according to their previous experience with ICT and to the results of a short assessment of basic ICT skills. Respondents can also simply refuse to take the test on a computer, in which case they are directed to the paper-based instruments. In the computer-based branch, respondents are assessed in (at most) two of three domains: literacy, numeracy, and problem solving in technology-rich environments. In the paper-based branch they first undertake a core assessment in literacy and numeracy, and then are randomly assigned to either a literacy or a numeracy assessment. Finally, they undertake the same reading component assessment that is used in STEP. Respondents who fail the core are only administered the reading components assessment.

Figure 10. The STEP assessment design



Source: World Bank (2014[8]), "STEP skills measurement surveys: innovative tools for assessing skills", http://documents.worldbank.org/curated/en/516741468178736065/STEP-skills-measurement-surveys-innovative-tools-for-assessing-skills (accessed on 1 April 2020).

STEP and PIAAC use similar psychometric methods to estimate the proficiency of participating adults. The scale on which literacy proficiency is evaluated is common across the two surveys, so that the results are directly comparable. Still, the differences in the underlying distribution of proficiency of the population may have undesirable consequences on the performance of the statistical methods used to estimate scores, especially when a large portion of the population perform at the very bottom of the proficiency distribution. These issues will be discussed in more details in Chapter 3.

3. Estimating the proficiency of low-skilled adults in PIAAC and STEP

The purpose of this chapter is to illustrate in more depth the instruments and the statistical tools used by PIAAC and STEP to estimate the proficiency of respondents. The differences between the two surveys will be highlighted, in order to discuss the extent to which they can influence a direct comparison of results from the two studies.

3.1. Proficiency is an unobservable latent trait

The PIAAC assessment was developed on the basis of a conceptual framework that defined literacy as "*understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential*" (PIAAC Literacy Expert Group, 2009[19]). This functional definition of literacy was substantiated in assessment items targeted to some extent to specific societies and to specific populations, those of adults living in OECD countries. There are important reasons for calibrating an assessment to a specific population. Contrary to physical attributes like age, income, or height, cognitive skills are not directly observable: they are a latent trait, which is (imperfectly) revealed through the answers given to assessment items. Studies like STEP, PIAAC, and other national and international studies, are based on the assumption that the "true" and unobservable proficiency of respondents lies somewhere in a unidimensional continuum. Statistical models developed under Item Response Theory (IRT) are used to estimate as best as possible where respondents lie in this continuum, on the basis of observed proxies such as the answers to a test.

An increase in the precision of the estimates of proficiency can normally be achieved by increasing the number of items administered in the assessment, as well as the range of item difficulty (where difficulty is defined as the level of latent proficiency needed to give a correct answer to an item). Assessments that are too easy or too difficult are not very informative, as they do not discriminate participants according to their performance (which would be uniformly very high in case of an easy assessment and uniformly very low in case of a difficult assessment). An assessment is most informative when it contains items the difficulty of which is close to the true underlying ability of the respondents, as this generates a smoother ranking of respondents in terms of percentage of correct answers, which can be expected to reflect the underlying ranking in terms of latent proficiency. In practice, assessments are restricted in terms of their possible length. This is one reason why PIAAC adopted an adaptive design, in which blocks of items of different difficulties were administered to different individuals according to their expected proficiency (e.g., more highly educated individuals, who are expected to have higher latent proficiency, had higher probabilities of answering more difficult items). This allows a reduction in the number of items administered to any single respondent and, therefore, a reduction in the length of the assessment. Furthermore, as large-scale assessments like PIAAC and STEP are mainly interested in the distribution of proficiency in the overall target population, rather than in the proficiency of any single individual, adjustments to the individual-level data are made to increase the precision of the estimates at the population level (Von Davier, Gonzalez and Mislevy, 2009[20]).This is done by coupling IRT models with a so-called population model, which uses background characteristics of respondents (such as age or educational attainment) to estimate more precisely the distribution of proficiency in the population. In addition, these models can also be used to estimate the proficiency of respondents in domains in which they were not actually tested. This is done in PIAAC, where all respondents receive a score in literacy, numeracy and problem solving in technology-rich environment, despite the fact that each participant undertakes an assessment in, at most, two domains. The score in the domain that the respondent did not attend is imputed on the basis of background characteristics, performance in the two other domains, and the relationship between background characteristics and proficiency in the entire sample.

A direct consequence of these features is that an assessment might not perform equally well across very heterogeneous settings or, to put it in a different way, that administering an assessment to populations with very different levels of proficiency would require a large pool of items in order to better cover the range of proficiency of respondents. Indeed, the suitability of PIAAC as a global tool to assess adult literacy has been often questioned (UNESCO, 2018[21]). The administration of STEP in low and middle-income countries, and the expansion of PIAAC (which in its second and third round was implemented in a wider range of countries, including some that are classified as middle-income) offer an opportunity to evaluate these issues on the basis of actual data.

Indeed, in some STEP and PIAAC countries a reasonably large share of the target population was not able to correctly answer the easiest literacy items. These tasks (classified as belonging to the proficiency level “below Level 1”) require the respondent to read brief texts on familiar topics to locate a single piece of specific information. There is seldom any competing information in the text and the requested information is identical in form to information in the question or directive. In case the respondent is required to locate information in short continuous texts, the information can be located as if the text was non-continuous in format. Only basic vocabulary knowledge is required, and the reader is not required to understand the structure of sentences or paragraphs or make use of other text features (OECD, 2016[22]). In one of such items, respondents are asked to use a notice providing results from a union election to identify the candidate with the fewest number of votes. Although the notice contains several paragraphs of information, the respondent only

needs to use a very short table with three numbers and associated names within the text to answer the question. The key word ("votes") appears in both the prompt and the text making the relevant information very transparent. There is no competing information as the word "votes" appears nowhere else in the text. To locate the answer, the respondent needs to compare the three numbers (the word "fewest" in the prompt indicates the answer will involve a number), and once that is determined, locate the name associated with that number (OECD, 2016[23]).

3.2. The core assessment in PIAAC and STEP

In both PIAAC and STEP, the easiest items (such as the one detailed above) were administered as part of an initial and short module of the assessment, the purpose of which was to identify respondents with very low proficiency who would have little chance of successfully completing most, if not all, items included in the assessment. Respondents who fail this short module, called the "core" assessment, did not continue with the survey and were not administered the main assessment.

The core assessments are not identical in the two studies. As STEP only assesses literacy, the STEP core contains 8 literacy items. In order to pass the core and move on to the full assessment, respondents must give a correct answer to at least three items. The PIAAC core assesses both literacy and numeracy. The paper-based version of the core (administered to respondents who fail a basic assessment of ICT skills or that refuse to be tested on a computer) contains four literacy and four numeracy items, and respondents must answer correctly at least four items. The four literacy items in the PIAAC core tests are also present in the STEP core test. The computer-based version of the core contains three literacy and three numeracy items, and respondents are required to answer correctly at least three items.

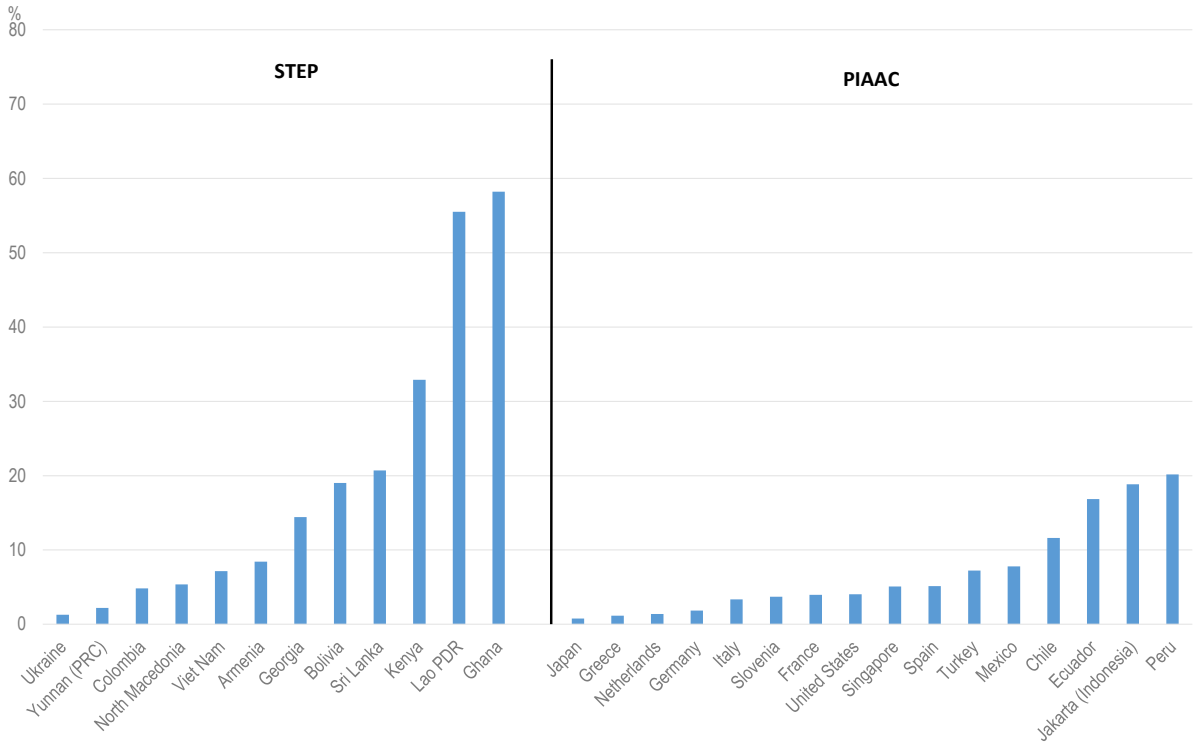
The reason for introducing this screening mechanism was mainly practical: failure to pass the core was considered as an indication that the skills of these adults were so low that undertaking the full assessment would have generated little additional information and, in addition, been an extremely frustrating and negative experience for the respondent. Conceptually, it is however not clear how to interpret the fact that a respondent failed the core. If adults who fail the core would not be able to answer correctly any item of the full assessment, it could actually be argued that this is due to a limitation of the assessment, which was not designed to assess proficiency at the lower end of the skills distribution (i.e. it did not contain any items of a sufficiently low level of difficulty to positively define the skills of these adults).

The scale on which PIAAC and STEP scores are measured is continuous and unbounded, which means that it is not correct to state that adults failing the core are "off the scale": indeed, they did receive a score, estimated using only a limited amount of background information from the BQ (such as age and educational attainment) and performance in the core.

The share of adults failing the core is generally higher in lower-income countries, which are over-represented in STEP (Figure 11)[2]; it varies widely, ranging from below 5% in countries like Ukraine and Germany, to around 10% in Chile and Georgia, up to almost 60% in Lao PDR and Ghana.

[2] For the sake of simplicity, Figure 11 shows results for only a subset of PIAAC countries and economies. In particular, the choice was made to include all upper-middle income countries (where proficiency is on average lower, and where a higher share of adults failed the core assessment), and a few higher-income countries. A similar criterion is adopted for all the Figures in Chapters 3, 4 and 5. As far as Figure 11 is concerned, the share of population failing the core assessment in the PIAAC countries and economies omitted from the graph is around or below 5%.

Figure 11. Share of the population failing the core assessment



Note: Except for Lao PDR and Sri Lanka, the data for STEP countries only refer to the adult population living in urban areas. PRC stands for People's Republic of China.
Sources: OECD ([24]), Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

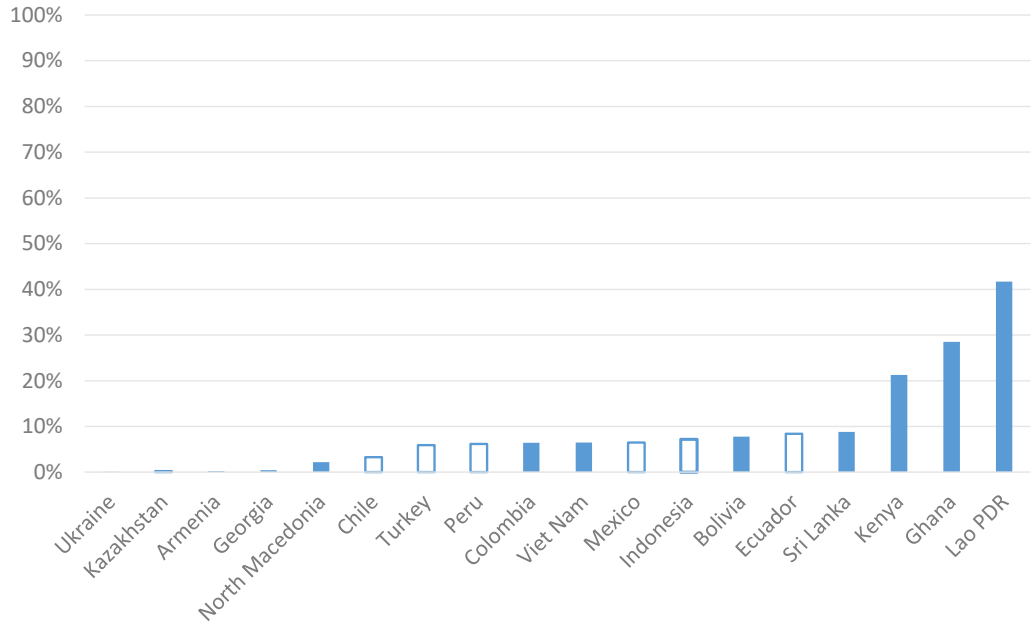
These people however can hardly be considered "illiterate", at least in the traditional sense of being able to read and write – more precisely, *the ability to read and write, with understanding, a short, simple statement about one's everyday life* (United Nations, 2008[25]). This is for instance the definition adopted by the UNESCO Institute of Statistics (UIS), and currently used to monitor progress towards the Sustainable Development Goal 4, which aims to ensure that by 2030 "all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy". While the UIS acknowledges that the definition of literacy has evolved over the decades, moving from a basic dichotomous variable (literate/illiterate) to a more nuanced definition of functional literacy and the foundation of lifelong learning, statistics on traditional (dichotomous) literacy rates are much more widely available than the national assessments required to produce more robust and detailed measures of literacy, as they are much easier and cheaper to collect: normally, they are collected during surveys or censuses, and are based on self-declaration or declaration by the designated respondent, who responds on behalf of all members in the household.

Assuming that these data do not suffer from severe under-reporting (which could arise in case adults were reluctant of admitting they are illiterate), the shares of adults that are considered illiterate are generally lower than the percentage of adults failing the core, as displayed in Figure 12. Moreover, it should be noted that the statistics on literacy rates refer to the entire country, and not to urban areas, as is the case for STEP data. Given the differences in educational attainment between urban and rural areas, it is reasonable to

assume that in the urban areas of countries that participated in STEP the share of illiterate adults would be even lower.

Figure 12. Share of illiterate adults

Adults aged 15 or above



Notes: Data refer to 2011, with the following exceptions (due to data availability): 2009 for Viet Nam, 2010 for Ghana, Kazakhstan and Sri Lanka, 2012 for Peru and Ukraine, 2014 for Georgia, Kenya and North Macedonia. Countries that participated in STEP are displayed in blue, while countries and economies that participated in PIAAC are displayed in white.
Source: UNESCO Institute for Statistics ([26]), *UIS.stat*, http://data.uis.unesco.org/ (accessed on 16 April 2020).

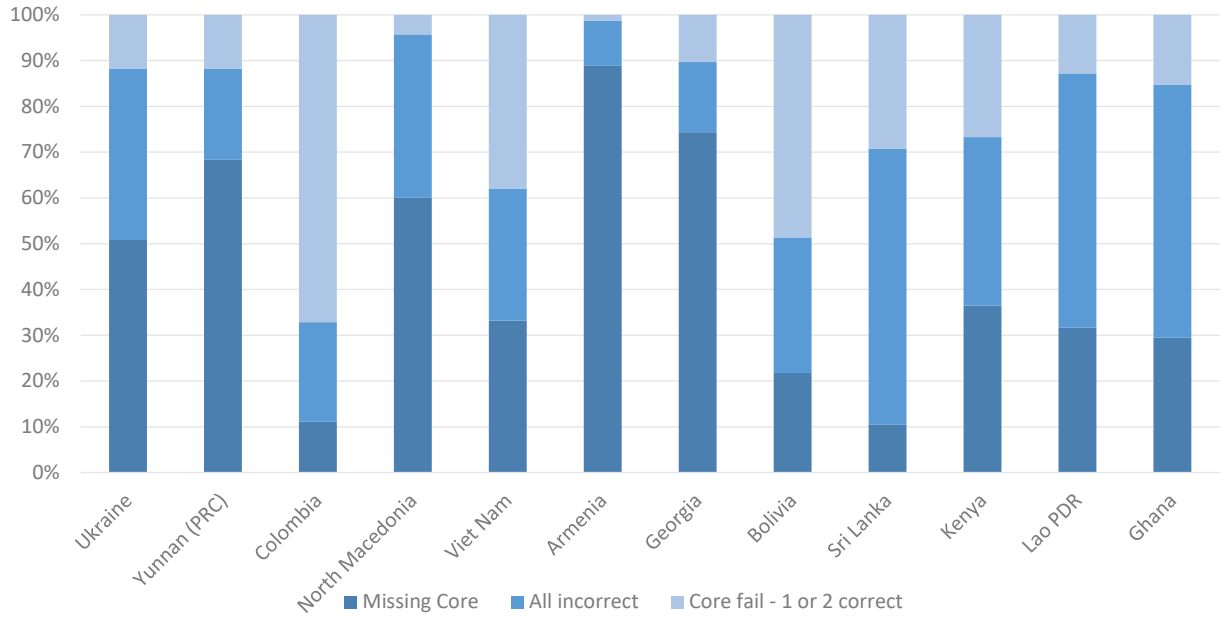
By the nature of the discrete rule that determines success and failure in the core, the group of respondents who fail the core includes individuals who get no items correct and those who get one or two items correct; there are also cases of respondents for whom information is missing, presumably because they did not even attempt the core. The distribution of respondents across these categories varies considerably between countries. Importantly, in some countries there is a reasonable large proportion of respondents for whom information regarding performance on the core items is missing.

3.3. Estimating the proficiency of respondents who failed the core

The most important implication of having a large share of respondents failing the core is that the estimates of the literacy proficiency of these adults rely on a small amount of information on actual performance, and much more on their background characteristics and on the design of the underlying statistical model used to estimate proficiency (the population model). The larger amount of error surrounding estimates at the lower end of the proficiency distribution is not a big issue for countries where the percentage of people failing the core is very low, because their noisily estimated scores would not have a large impact on the estimated distribution of proficiency at the population level. This was the case for the large majority of PIAAC countries. When a large number of adults have their score imputed through a model-based procedure, this can lead to more imprecise estimation

of the entire distribution of proficiency of the target population, as will be discussed in more details in the rest of this chapter.

Figure 13. Reasons for failing the core in STEP countries

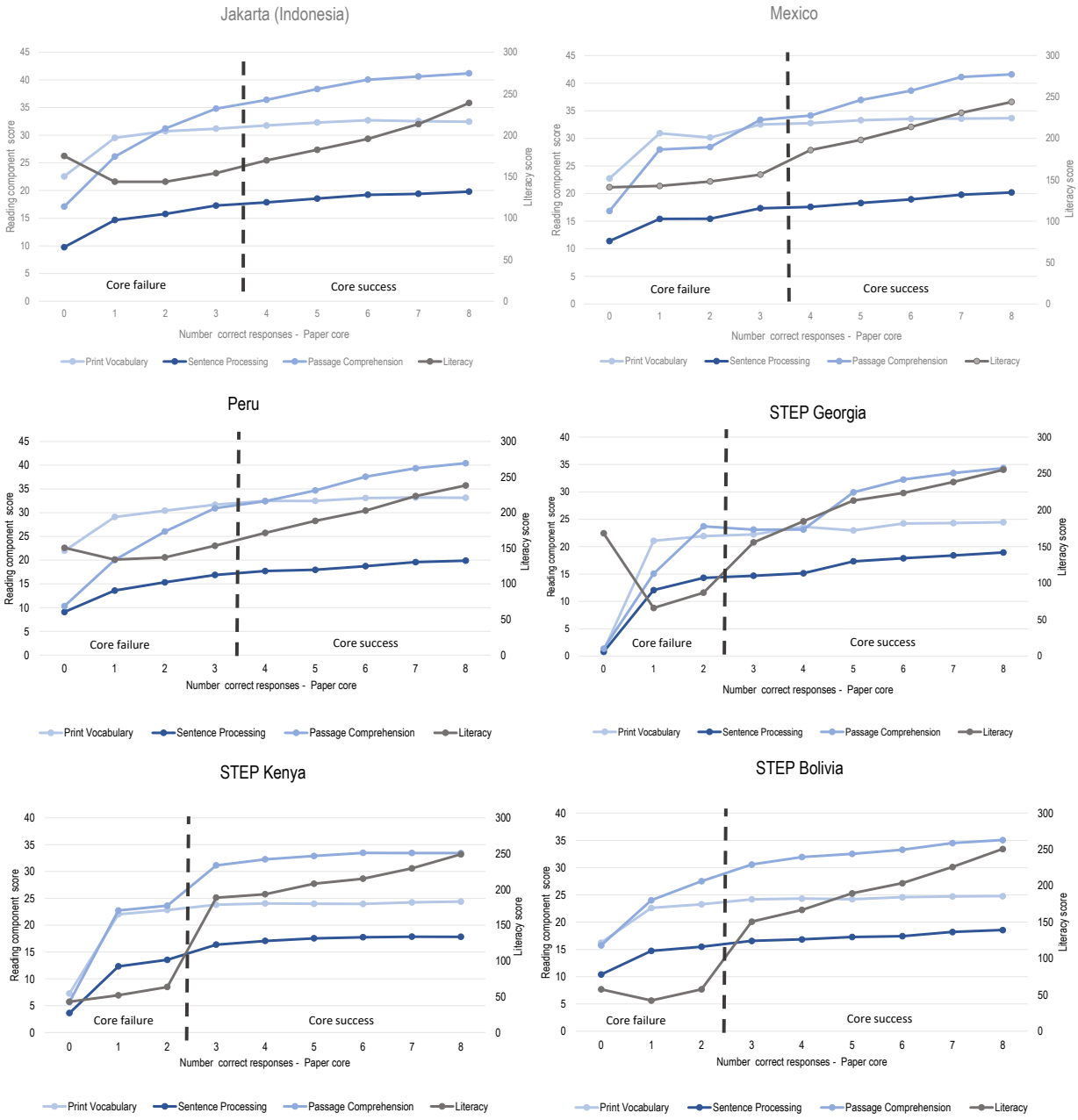


Note: Except for Lao PDR and Sri Lanka, the data for STEP countries only refer to the adult population living in urban areas. PRC stands for People's Republic of China. Countries are sorted by the overall percentage of respondents failing the core.
Source: World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

3.4. The reading components assessment

Some additional information on what adults at the bottom end of the proficiency distribution are actually able to do can be found in the reading components assessment. The reading components were however not designed to measure the concept of literacy, as defined in the PIAAC assessment framework; moreover, results in the reading components cannot be directly compared between PIAAC and STEP. STEP administered the reading component assessment to all participating adults (and therefore also to those with high literacy proficiency), while PIAAC administered it only to adults who failed the core.

Figure 14. Performance in the core and in the reading components



Note: The figure plots scores in reading components (left-hand scale of the vertical axis) and in literacy (right-hand scale of the vertical axis) against the number of correct answers in the paper core (horizontal axis). The data for STEP countries only refer to the adult population living in urban areas.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Reading components can still be useful to evaluate the informational content of performance in the core, as one could compare the number of correct answers in the core with performance in the reading component assessment. Figure 14 does that for some selected STEP countries and for those PIAAC countries where a significant proportion of adults failed the core. In all countries, performance in the reading components increases more or less in a linear relationship with the number of correct answers in the core.

This is not always the case for the estimated literacy score. In Georgia, Jakarta (Indonesia), and Peru, the estimated average score of adults who did not give any correct answer to the core is higher than the score estimated for adults answering 1, 2 or even 3 tasks correctly. This suggests that there is a large measurement error associated with the estimated score for respondents at the very bottom of the skill distribution.

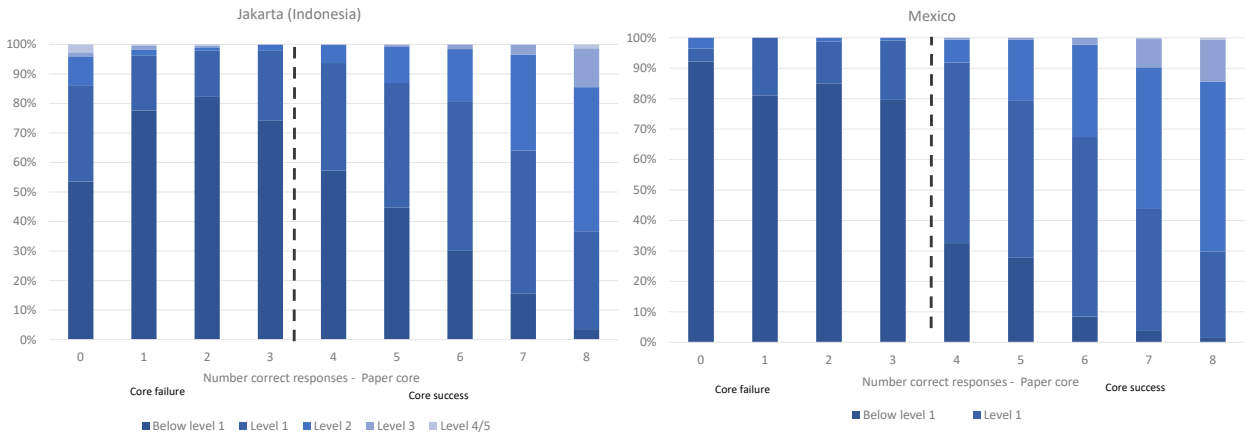
Importantly, a discontinuity is apparent in the relationship between performance in the core and estimated literacy proficiency, due to the fact that a discrete threshold determines success or failure in the core, and therefore whether or not the literacy score is mainly imputed from background information or is based on actual performance in the full literacy assessment. This constitutes an indirect test of how well the underlying model is able to estimate latent proficiency: if one is willing to assume that respondents who only just failed the core and respondents who only just pass the core should have similar levels of proficiency, and that estimated literacy proficiency should correlate positively with the number of correct answers given in the core, one would expect literacy proficiency to be approximately a linear function of the number of correct answers in the core (at least locally, i.e. in a neighbourhood of the threshold that determines success in the core).

3.5. Implications for the estimation of proficiency at the bottom end of the distribution

A first implication of these results is that researchers should give preference to analysing proficiency levels, rather than continuous proficiency scores, as one would expect the model to at least being able to correctly classify respondents according to broad categories. This is shown in Figure 15 for two PIAAC countries/economies (Mexico and Jakarta [Indonesia]) and in Figure 16 for two STEP countries (Ghana and Bolivia).

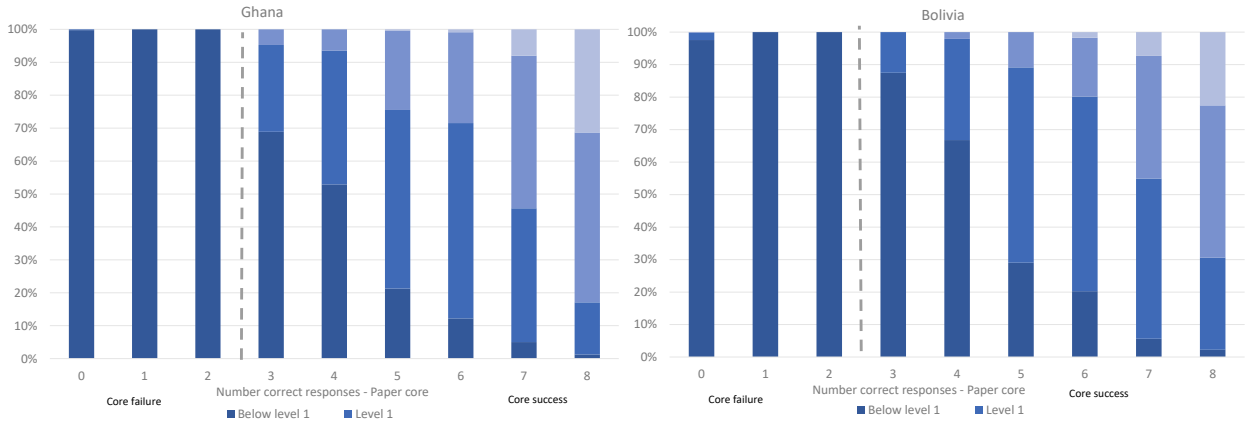
In all countries, the share of adults at the different proficiency levels changes in a regular way for people who passed the core see the results presented on the right of the dashed line in Figures 15 and 16): proficiency for these individuals is precisely estimated on the basis of the full assessment. This is not always the case (see the results on the left of the dashed line). In Bolivia and Ghana, all individuals who failed the core are classified at 'below Level 1' in the proficiency scale, even if they answered correctly to 2 questions of the core. In Jakarta (Indonesia), adults who gave 2 or 3 correct answers to the core are more likely to be classified as 'below Level 1' than those who gave 0 correct answers. The profile is smoother and better-behaved in Mexico, but also in that case there is a large discontinuity in the probability of being classified 'below Level 1' between people when moving from 3 to 4 correct answers in the core.

Figure 15. Levels of literacy proficiency and performance in the core in Jakarta (Indonesia) and Mexico



Note: The figure plots the percentage of adults scoring at different proficiency levels (on the vertical axis) for different number of correct answers given in the core assessment (on the horizontal axis).
Source: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020).

Figure 16. Levels of literacy proficiency and performance in the core in Ghana and Bolivia



Note: The figure plots the percentage of adults scoring at different proficiency levels (on the vertical axis) for different numbers of correct answers given in the core assessment (on the horizontal axis). The data for STEP countries only refer to the adult population living in urban areas.
Source: World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Table 6 summarises the results of this analysis by showing the average scores for respondents who passed the core, who failed the core, and who passed the core with at most 6 correct answers out of 8. In the selected PIAAC countries shown in the table, where a significant proportion of adults failed the core, the differences between those who failed and those who barely passed the core ranges between 30 and 50 score points. In STEP countries the differences are much larger, reaching 146 score points in Kenya. Armenia and Georgia are exceptions in this respect; in Georgia respondents who failed the core have actually *higher* average scores than respondents who passed with 4 to 6 correct answers. This seems to suggest that the models used to estimate the scores of respondents who failed the core are different, or at least work differently, in PIAAC and STEP, with either the

model used in PIAAC overestimating the scores, or the model used in STEP underestimating them. These models are based on both the results of the assessment and on information on background characteristics collected in the questionnaire and that are expected to be correlated with proficiency. In the case of respondents for whom most information from the assessment is mostly missing (by design), the estimation of scores will rely primarily on background characteristics and will, therefore be associated with a high degree of error. The higher the proportion of adults failing the core, the more the estimated scores will rely on the use of background information, rather than test performance.

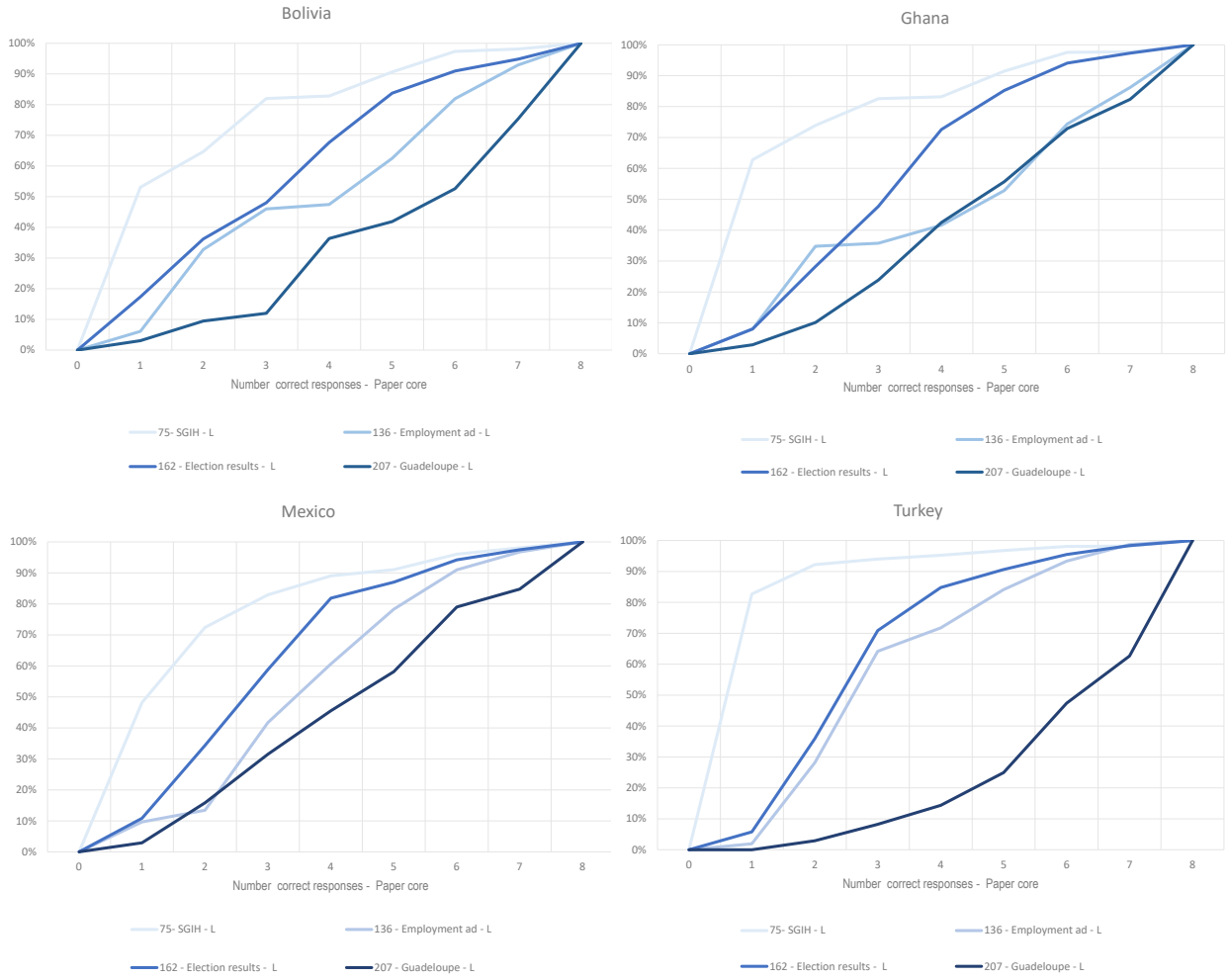
Table 6. Performance in the core and estimated literacy score

	Average score for respondents who passed the core (1)	Average score for respondents who failed the core (2)	Average score for respondents with 4-6 correct answers in the core (3)	Difference column 3-column 2
STEP countries				
Armenia	260	216	232	16
Bolivia	222	101	187	87
Colombia	245	81	197	116
Georgia	249	228	209	-19
Ghana	221	75	193	118
Kenya	228	58	204	146
Ukraine	270	100	220	120
Viet Nam	253	108	198	90
PIAAC countries				
Chile	230	161	198	37
Ecuador	209	141	191	50
Mexico	232	156	204	48
Peru	214	145	187	41
Turkey	234	174	206	31

Note: The data for STEP countries only refer to the adult population living in urban areas. The difference between column 3 and column 2 may not be precise because of rounding.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

The differences in the estimated proficiency of respondents who failed the core really seem to be attributable to the imputation model rather than to the differential functioning of items in PIAAC and STEP countries. Figure 17 shows response curves for common core items, i.e. the four items that are present in both the PIAAC and STEP core. These graphs plot the probability of giving a correct answer to a given item (on the vertical axis) against the total number of correct answers in the core. Evidently, people who gave 8 correct answers to the core have a 100% probability of giving a correct answer to any core item, because they answered correctly all 8 items in the core; similarly, people who gave 0 correct answers have a 0% probability of giving a correct answer to any core items. The differences in the shape of the various curves is due to the varying difficulty of each item: the probability of answering correctly to the Guadeloupe A item (which has a difficulty of 207) is consistently lower than the probability of answering correctly an easy item like SIGH (whose level of difficulty is 75). The important message from Figure 17 is that the response curves are much the same in all countries, whether they participated in STEP (like Bolivia and Ghana) or in PIAAC (Mexico and Turkey).

Figure 17. Response curves for common core items



Note: The graph displays the percentage of respondents who answered correctly to the item (on the vertical axis) against the overall number of correct responses in the core (on the horizontal axis). The data for STEP countries only refer to the adult population living in urban areas.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

3.6. Conclusions

Cognitive assessments are normally designed to give precise estimates of the underlying proficiency towards the middle of the distribution, where the majority of the target population is concentrated, accepting a larger degree of estimation error at the tails of the distribution: scores at the top and at the bottom ends are normally noisier, because the assessments contain a limited number of very easy or very difficult items.

This is normally not a big problem for large-scale assessments whose aim is to measure the distribution of proficiency in the population, as long as assessment items are well targeted to the proficiency of a large portion of the population.

In some PIAAC and STEP countries, however, a significant share of the population was found to be unable to answer even the easiest items in the assessment. The presence of a core assessment, which screens adults at the low end of the distribution out of the full

assessment, exacerbates the problem, because the scores of a significant number of respondents are estimated on the basis of a very limited number of items, complemented with information on their background information; as a result, the estimated distribution of the proficiency of the population relies to a greater extent on the statistical model used to estimate proficiency. These weaknesses can be somehow attenuated by analysing more robust indicators such as the percentage of people at different levels of proficiency.

More in-depth analysis reveals that performance in the reading components is correlated with performance in the core, but also that the statistical models used in PIAAC and STEP tend to deliver different estimates of proficiency for adults with similar performance in the core.

4. Literacy proficiency in PIAAC and STEP countries

Bearing in mind the caveats regarding the comparison of the results of the two surveys discussed in previous chapters (and summarised in Box 2), this chapter compares the results of STEP and PIAAC regarding the literacy proficiency of adults. The chapter will present results for all STEP countries and for a number of selected PIAAC countries (focussing in particular on upper-middle income countries, which are more similar to STEP countries in terms of level of economic development and estimated adults' proficiency).

Box 2. Differences between PIAAC and STEP

The literacy assessment administered in STEP is based on the paper-based assessment of literacy used in PIAAC. From a psychometric point of view, the two assessments were designed to be entirely comparable, as they are based on the same instruments and on similar psychometric models, which produce scores on a common scale.

There are however differences in the design of the survey and in other methodological aspects that prevent full comparability between the results of the two studies. These issues have been discussed at length in Chapters 2 and 3, and can be summarised as follows:

- The target population in STEP is limited to adults living in urban areas, while PIAAC is representative of all adults living in a country (or in a sub-national entity)
- STEP surveys adults aged 15-64, while PIAAC surveys adults aged 16-65
- STEP uses only paper-based instruments, while the PIAAC assessment was designed to be primarily administered on a computer (although respondents had the possibility of opting out of the computer-based assessment and use paper-based instruments)
- The rules governing success on the core assessment (a short test used to screen low-proficiency adults out of the full assessment) are different: respondents in STEP had to successfully complete 3 literacy items out of 8, while in PIAAC they have to successfully complete at least 4 items (of any domain) out of a pool made of 4 literacy items and 4 numeracy items.

Because of the extent of these differences, and in particular the fact that the sampled units are not directly comparable, this chapter will be mostly focused on the analysis of the relationship between literacy skills and socio-demographic characteristics of respondents, as this can provide insights into the validity of the assessment instruments and the appropriateness of the PIAAC assessment to serve as a basis for a common instrument to assess adult skills on an (almost) global scale.

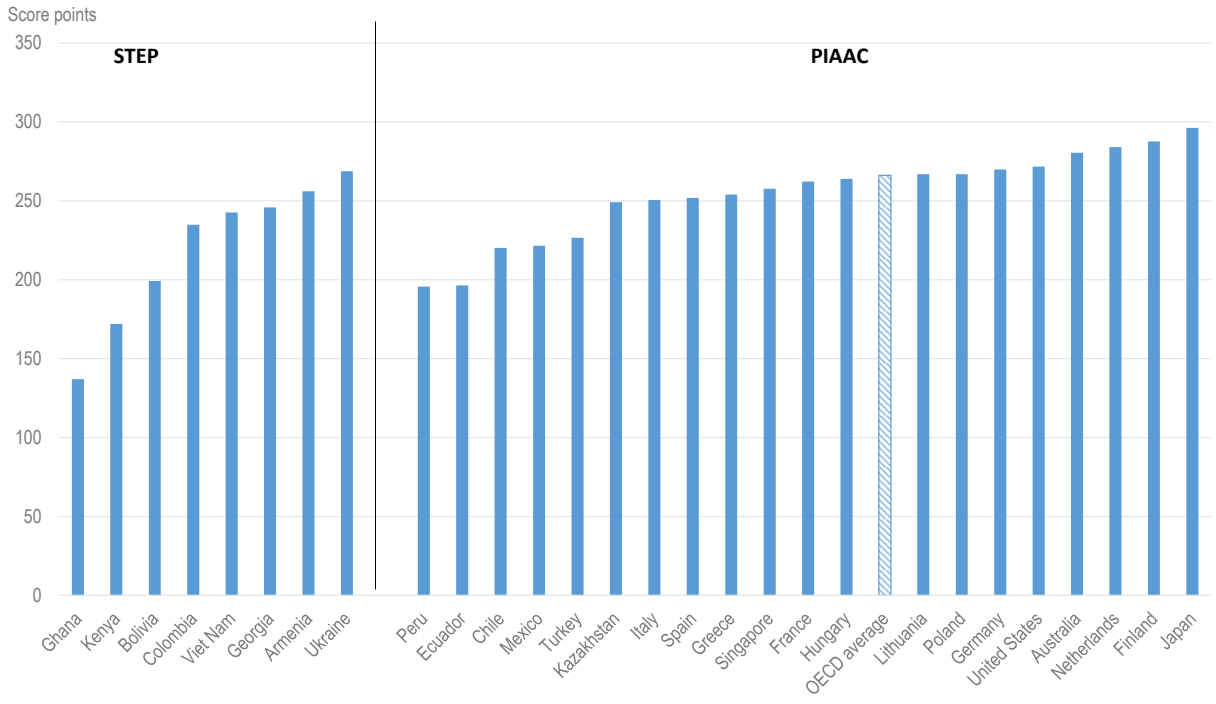
The differences in the sampling schemes between STEP and PIAAC imply that making reference to 'countries' when discussing the results of STEP represents something of a

shorthand. The results from STEP cover only a subset of the national population in the countries participating in the study – the adult population in 'urban areas', the exact definition and size of which vary by country, while the target population in PIAAC is the entire adult population residing in the national territory (or in a sub-national entity, which is defined on the basis of administrative boundaries rather than on characteristics such as the degree of urbanisation).

4.1. Literacy proficiency in PIAAC and STEP

Figure 18 displays average scores on the literacy assessment administered in PIAAC and STEP. For the reasons discussed at length in Chapter 2, the results are not directly comparable as country-wide differences in adults' proficiency, as the studies targeted slightly different populations, with STEP only assessing adults living in urban areas. For this reason, the countries are visually separated in all graphs.

Figure 18. Average literacy proficiency in PIAAC and STEP countries



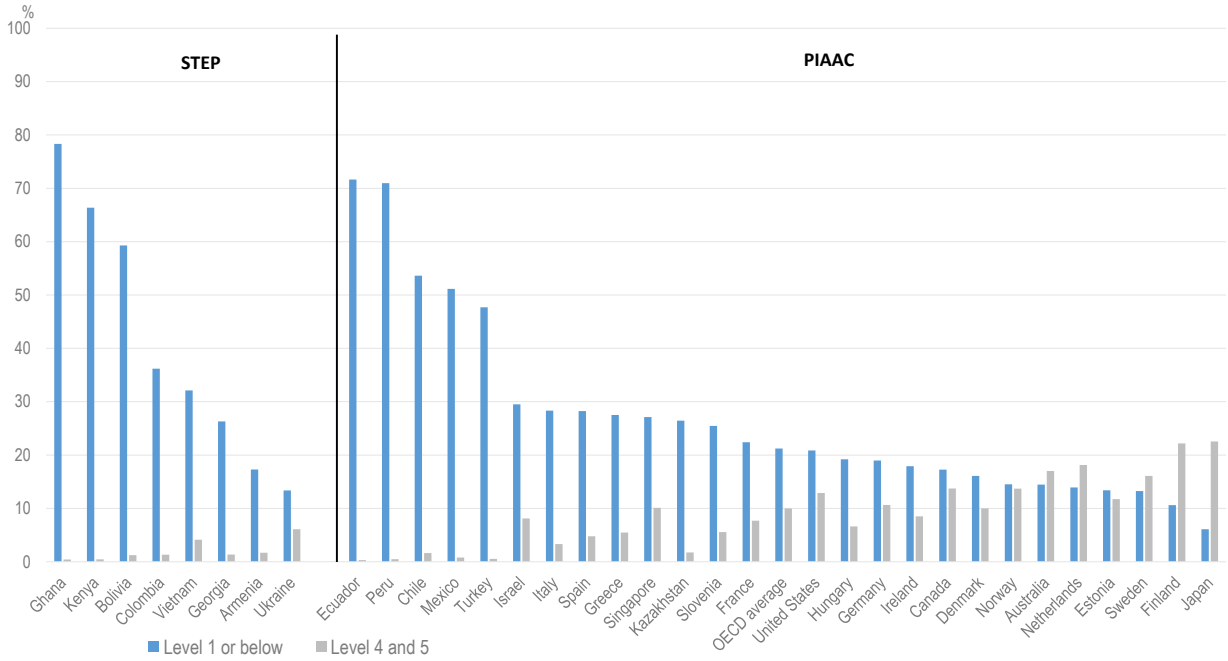
Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Taking into account the fact that the results for STEP countries cover adults in urban areas only, estimated average proficiency in literacy is largely in line with expectations. Average proficiency is extremely low in Ghana and Kenya compared to the OECD average. In the other STEP countries, the results are comparable to those in PIAAC countries in similar regions (eg. the former Soviet republics of Armenia, Georgia and Ukraine compared to Kazakhstan; and for Latin American countries: Bolivia and Colombia compared to Chile, Ecuador, Mexico and Peru). The large degree of variation in average literacy proficiency between STEP countries is notable: average literacy proficiency in the highest performing country in STEP (Ukraine) is at 269 points with a gap of about 130 score points with

respect to Ghana, which is the lowest performing country. This is far wider than the 100 score points that separate the highest and the lowest performing countries and economies in PIAAC (adults in Japan scored 296 points on average, while adults in Ecuador scored 196 points).

As argued in Chapter 3, in some countries, average scores can be influenced by the large degree of measurement error in the estimated scores of respondents who failed the core. For this reason, Figure 19 presents the percentage of respondents who scored at Level 1 or below and at Level 4 or 5. More details on what adults are able to do at different levels of proficiency are explained in Box 3.

Figure 19. Percentage of adults at different levels of proficiency



Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Figure 19 also gives an idea of the inequalities within countries in how literacy skills are distributed. The share of adults scoring at Level 1 or below is extremely high (above 70%) in Ghana, Ecuador and Peru, and is above 50% in Kenya, Bolivia and Chile. This is not only due to adults failing the core. Apart from Ghana (where almost 60% of adults failed the core) and Kenya (where about 30% of adults failed the core), the other countries from South America only had between 10 and 20% of respondents who failed the core assessment. In Armenia and Ukraine the share of adults at the two lowest levels of the literacy scale is about 15%, which is below the OECD average of 21%. However, in none of the countries that participated in STEP the share of adults scoring at Level 4 and 5 approached the levels observed in the average OECD country that participated in PIAAC (10%): while in Ukraine 6% of adults reached the two highest levels (a value similar to those observed in other Eastern European countries like Hungary, Lithuania and the Slovak Republic), this share is much lower in all the other STEP countries.

Box 3. The proficiency levels

The proficiency scale used in both PIAAC and STEP can be described in relation to the items that are located at the different points on the scale according to their difficulty. To help interpret the results, the reporting scales have been divided into "proficiency levels" defined by particular score-point ranges. Six proficiency levels are defined for literacy (Levels 1 through 5 plus below Level 1). These descriptors provide a summary of the characteristics of the types of tasks that can be successfully completed by adults with proficiency scores in a particular range. In other words, they offer a summary of what adults with particular proficiency scores can do. With the exception of the lowest level (below Level 1), tasks located at a particular level can be successfully completed approximately 50% of the time by a person with a proficiency score at the bottom of the range defining the level. In other words, a person with a score at the bottom of Level 2 would score close to 50% in a test made up of items of Level 2 difficulty. A person at the top of the level will get items located at that level correct most of the time. The "average" individual with a proficiency score in the range defining a level will successfully complete items located at that level approximately two-thirds of the time.

The following table provides a more concrete description of the tasks that respondents at different proficiency levels are able to solve:

Level	Score range	Types of tasks completed successfully at each level of proficiency
Below Level 1	Lower than 176	The tasks at this level require the respondent to read brief texts on familiar topics to locate a single piece of specific information. There is seldom any competing information in the text and the requested information is identical in form to information in the question or directive. The respondent may be required to locate information in short continuous texts. However, in this case, the information can be located as if the text were non-continuous in format. Only basic vocabulary knowledge is required, and the reader is not required to understand the structure of sentences or paragraphs or make use of other text features. Tasks below Level 1 do not make use of any features specific to digital texts.
1	176-225	Most of the tasks at this level require the respondent to read relatively short digital or print continuous, non-continuous, or mixed texts to locate a single piece of information that is identical to or synonymous with the information given in the question or directive. Some tasks, such as those involving non-continuous texts, may require the respondent to enter personal information onto a document. Little, if any, competing information is present. Some tasks may require simple cycling through more than one piece of information. Knowledge and skill in recognising basic vocabulary determining the meaning of sentences, and reading paragraphs of text is expected.
2	226-275	At this level, the medium of texts may be digital or printed, and texts may comprise continuous, non-continuous, or mixed types. Tasks at this level require respondents to make matches between the text and information, and may require paraphrasing or low-level inferences. Some competing pieces of information may be present. Some tasks require the respondent to: • cycle through or integrate two or more pieces of information based on criteria; • compare and contrast or reason about information requested in the question; or • navigate within digital texts to access and identify information from various parts of a document.
3	276-325	Texts at this level are often dense or lengthy, and include continuous, non-continuous, mixed, or multiple pages of text. Understanding text and rhetorical structures become more central to successfully completing tasks, especially navigating complex digital texts. Tasks require the respondent to identify, interpret, or evaluate one or more pieces of information, and often require varying levels of inference. Many tasks require the respondent to construct meaning across larger chunks of text or perform multi-step operations in order to identify and formulate responses. Often tasks also demand that the respondent disregard irrelevant or inappropriate content to answer accurately. Competing information is often present, but it is not more prominent than the correct information.
4	326-375	Tasks at this level often require respondents to perform multiple-step operations to integrate, interpret, or synthesise information from complex or lengthy continuous, non-continuous, mixed, or multiple type texts. Complex inferences and application of background knowledge may be needed to perform the task successfully. Many tasks require identifying and understanding one or more specific, non-central idea(s) in the text in order to interpret or evaluate subtle evidence-claim or persuasive discourse relationships. Conditional information is frequently present in tasks at this level and must be taken into consideration by the respondent. Competing information is present and sometimes seemingly as prominent as correct information.

5	Higher than 376	At this level, tasks may require the respondent to search for and integrate information across multiple, dense texts; construct syntheses of similar and contrasting ideas or points of view; or evaluate evidence-based arguments. Application and evaluation of logical and conceptual models of ideas may be required to accomplish tasks. Evaluating reliability of evidentiary sources and selecting key information is frequently a requirement. Tasks often require respondents to be aware of subtle, rhetorical cues and to make high-level inferences or use specialised background knowledge.

Source: OECD (2016[22]), *The Survey of Adult Skills: Reader's Companion, Second Edition*, https://dx.doi.org/10.1787/9789264258075-en.

4.2. Literacy proficiency and educational attainment

A robust result of the analysis undertaken so far with PIAAC data is the strong correlation between skills proficiency and educational attainment. This result also provides evidence of the validity of the PIAAC instruments, as they measure something that correlates as expected with educational attainment. On average across OECD countries and economies that participated in PIAAC, tertiary-educated adults score 60.5 points higher than adults without upper secondary education; the gap decreases slightly (to 48 score points) after accounting for differences in other background characteristics like gender, age, immigrant and language background, and parents' educational attainment (OECD, 2019[27]).

A similar relationship is present in countries that participated in STEP, as shown in Table 7. Restricting the analysis to adults aged above 25 (in order to account for the fact that younger adults may still be in education), average literacy proficiency increases with educational attainment in all countries. Among STEP countries, the differences between tertiary-educated adults and adults who have not attained upper secondary are particularly large in Ghana and Kenya, and smallest in Armenia and Georgia.

Table 7. Average scores in literacy, by educational attainment

Adults aged 25-65

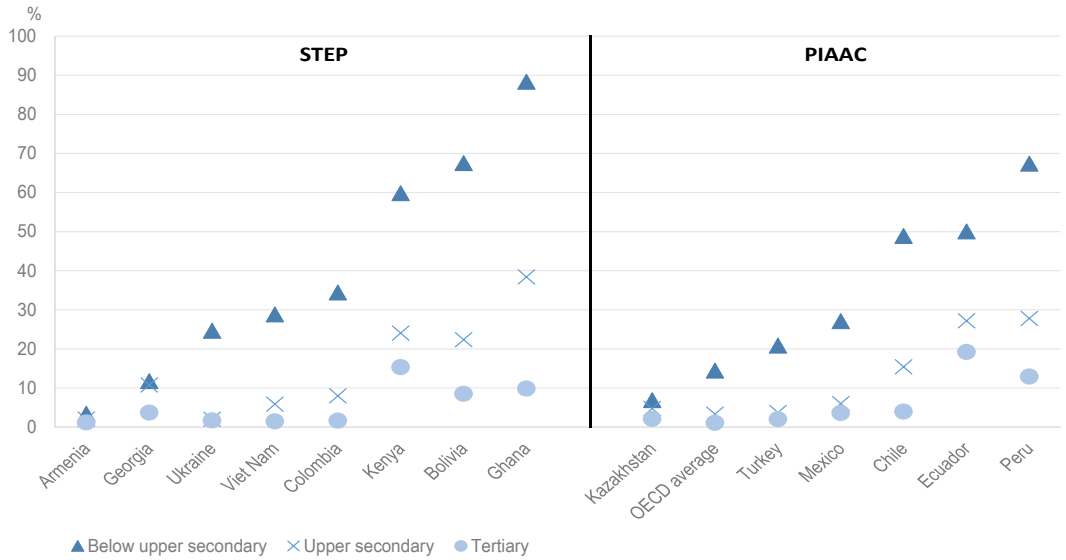
	Lower than upper secondary		Upper secondary		Tertiary	
	Mean score	S.E.	Mean score	S.E.	Mean score	S.E.
STEP						
Armenia	241.1	(4.9)	247.7	(2.2)	260.4	(2.0)
Bolivia	123.2	(6.5)	206.0	(4.4)	235.4	(4.3)
Colombia	187.8	(5.5)	240.0	(3.6)	269.2	(2.9)
Georgia	226.3	(5.5)	230.5	(3.2)	247.3	(1.6)
Ghana	86.8	(3.4)	178.5	(5.4)	238.0	(4.9)
Kenya	123.6	(3.9)	199.7	(3.8)	241.3	(7.3)
Ukraine	209.2	(16.3)	264.4	(2.4)	273.6	(3.4)
Viet Nam	196.7	(3.1)	248.4	(2.6)	273.5	(2.0)
PIAAC						
Chile	176.9	(1.8)	219.0	(1.7)	254.0	(2.6)
Ecuador	174.0	(1.5)	202.5	(1.7)	219.2	(3.0)
Kazakhstan	235.8	(2.6)	245.5	(1.2)	257.7	(1.6)
Mexico	201.3	(1.3)	237.0	(1.9)	255.6	(2.3)
Peru	156.6	(2.0)	200.7	(1.3)	227.4	(1.4)
Turkey	210.3	(1.7)	244.6	(1.6)	258.3	(1.6)
OECD average	**230.3**	**(0.4)**	**263.0**	**(0.2)**	**290.8**	**(0.3)**

Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Similarly, higher levels of educational attainment are associated with a strong reduction in the likelihood that a respondent performs at low levels of proficiency, and by an increase in the likelihood of scoring at higher levels, as illustrated in Figure 20 and Figure 21.

Figure 20. Percentage of adults scoring below Level 1, by educational attainment

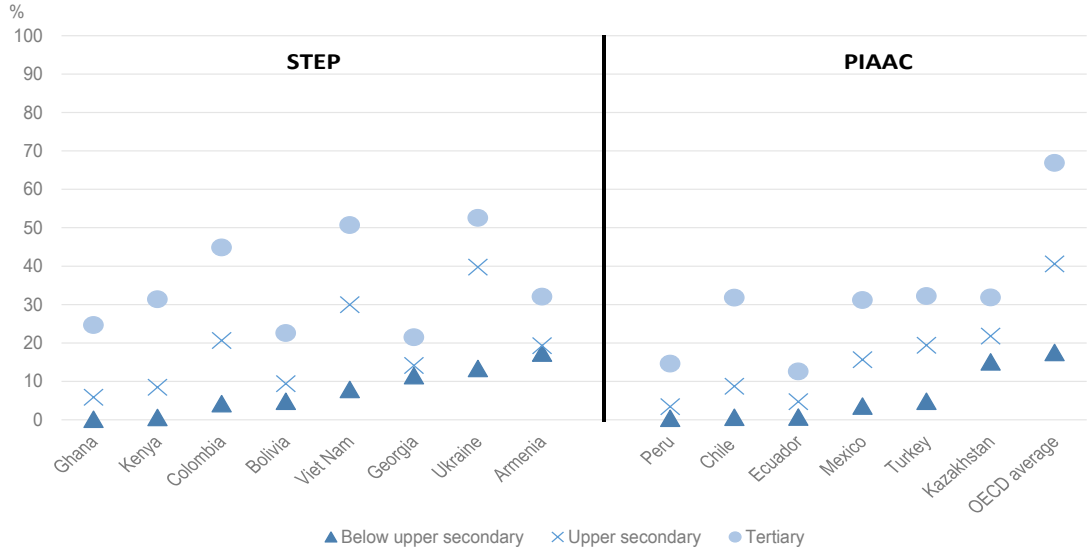
Adults aged 25-65



Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Figure 21. Percentage of adults scoring at Level 3 or above, by educational attainment

Adults aged 25-65



Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

4.3. Literacy proficiency and age

A second robust finding from large-scale international assessment of adult skills is the fact that the relationship between age and proficiency has an inverted-U shape, peaking for young adults in the mid-30s and then gradually declining as people get older. While this stylised fact holds in many countries, the relationship between age and gender in cross-sectional studies like PIAAC varies to a great extent across countries, as it depends heavily on the timing of the development of mass education in different countries, as well as on changes in educational quality over time. In other words, adults belonging to different birth cohorts have acquired different levels of education in different countries, and these so-called "cohort effects" influence the observed differences in proficiency across age groups at a given point in time.

Table 8. Average literacy proficiency, by age

	16-24 year-olds		25-34 year-olds		35-44 year-olds		45-54 year-olds		55-65 year-olds	
	Mean score	S.E.	Mean score	S.E.	Mean score	S.E.	Mean score	S.E.	Mean score	S.E.
STEP										
Armenia	260.1	(2.3)	257.1	(2.2)	258.0	(2.5)	253.0	(2.6)	248.8	(2.6)
Bolivia	224.8	(3.0)	210.5	(4.7)	190.1	(5.4)	162.5	(7.6)	150.2	(7.9)
Colombia	256.7	(4.0)	247.1	(3.8)	241.8	(4.2)	207.9	(8.4)	201.1	(6.8)
Georgia	258.8	(2.4)	244.8	(2.6)	247.1	(2.5)	240.5	(2.7)	235.4	(2.9)
Ghana	167.1	(4.8)	134.2	(4.2)	112.4	(5.5)	124.1	(7.4)	130.8	(7.1)
Kenya	182.6	(3.3)	173.4	(3.9)	162.8	(4.6)	154.2	(6.4)	127.0	(10.2)
Ukraine	271.0	(3.2)	272.5	(2.6)	261.5	(6.1)	268.1	(3.3)	266.1	(3.4)
Viet Nam	277.4	(2.1)	248.2	(3.6)	235.2	(3.5)	221.9	(3.5)	223.4	(4.0)
PIAAC										
Chile	237.0	(2.9)	235.3	(3.3)	220.4	(3.3)	206.2	(3.3)	193.9	(3.3)
Ecuador	205.9	(1.8)	202.5	(2.0)	192.4	(2.0)	189.4	(2.4)	181.0	(2.6)
Kazakhstan	248.9	(1.9)	249.1	(1.7)	250.8	(2.0)	247.6	(1.7)	249.0	(1.9)
Mexico	233.2	(1.5)	230.4	(1.8)	221.5	(2.0)	210.5	(2.1)	197.3	(2.4)
Peru	212.8	(1.7)	202.8	(1.8)	188.1	(2.2)	180.1	(2.3)	174.8	(2.3)
Turkey	236.6	(2.2)	234.0	(2.6)	225.1	(1.7)	221.5	(2.4)	204.3	(3.0)
OECD average	**273.5**	**(0.3)**	**277.0**	**(0.3)**	**271.2**	**(0.3)**	**260.9**	**(0.3)**	**248.4**	**(0.3)**

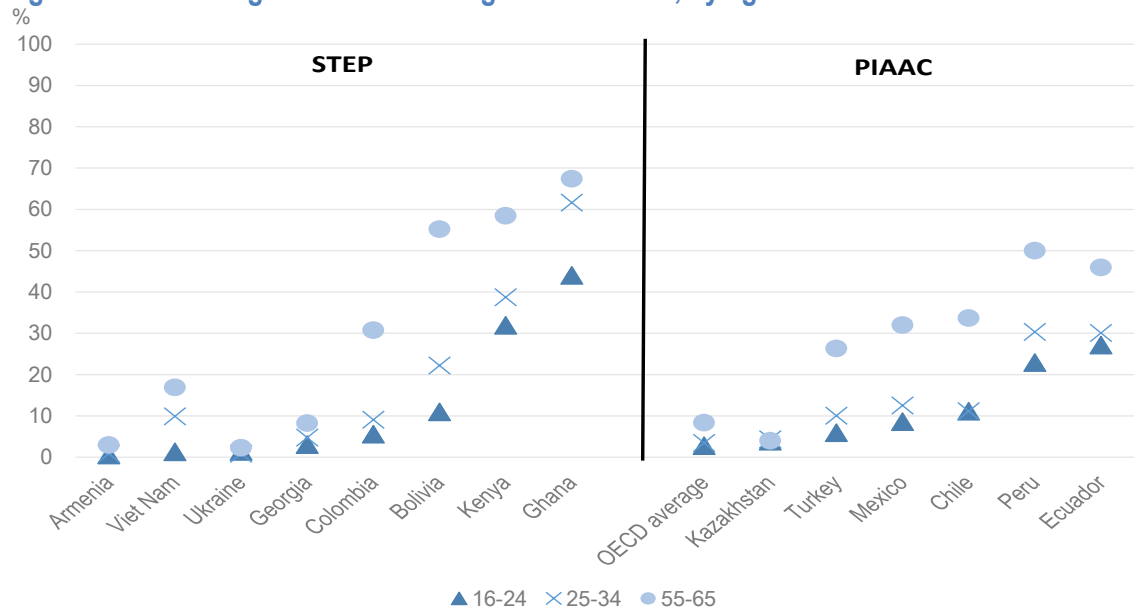
Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

On average across OECD countries and economies that participated in PIAAC, literacy proficiency is highest amongst 25-34 year-olds, and lowest among 55-65 year-olds, with an average difference between these two groups of almost 30 score points (Table 8).

However, in all STEP countries (with the exception of Ukraine), as well as in all Latin American countries that have participated in PIAAC, young adults aged 16-24 score on average marginally higher than adults aged 25-34, which is probably a sign that younger cohorts are acquiring more and/or better education. In most countries (with the exception of Ghana, Kazakhstan and Viet Nam), average proficiency is lowest among old adults aged 55-65, although the magnitude of the differences with respect to younger cohorts varies greatly across countries, being particularly high in Bolivia, Colombia, Kenya and Viet Nam, and much lower in Armenia, Kazakhstan and Ukraine.

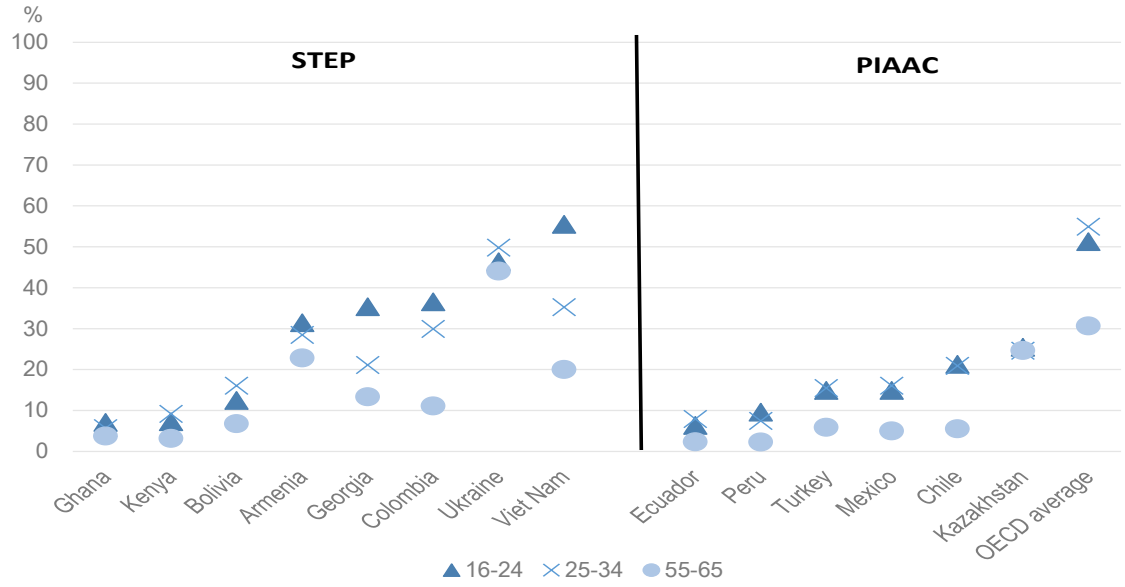
Figure 22 and Figure 23 show that older adults aged 55-65 have the highest probability of scoring below Level 1 and the lowest probability of scoring at Level 3 or above. In all countries analysed here, young adults aged 16-24 have the lowest chances of scoring below Level 1.

Figure 22. Percentage of adults scoring below Level 1, by age



Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Figure 23. Percentage of adults scoring at Level 3 or above, by age



Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

4.4. Literacy proficiency and gender

Gender differences in literacy proficiency in the adult population are normally very small: across OECD countries and economies that participated in PIAAC men scored on average only two points higher than women, a difference that, while statistically significant, can be considered irrelevant for practical purposes. Gender gaps are small in all PIAAC participating countries and economies, being highest in Turkey, where men have a small advantage of 11 score points.

Among countries that participated in STEP, gender differences are more pronounced. In Ghana the gap is extremely large, with men scoring 41 points higher than women on average; in other countries like Bolivia, Colombia, Kenya and Viet Nam, the gap is smaller (ranging between 7 and 16 score points). To put these gaps in perspective, the international standard deviation of literacy scores in PIAAC countries and economies is slightly below 50 score points; on average across countries and economies participating in PIAAC, an additional year of education is associated with an increase in literacy proficiency of around 7 points. In Georgia and Ukraine women score slightly higher than men, although in the latter the difference is not statistically significant.

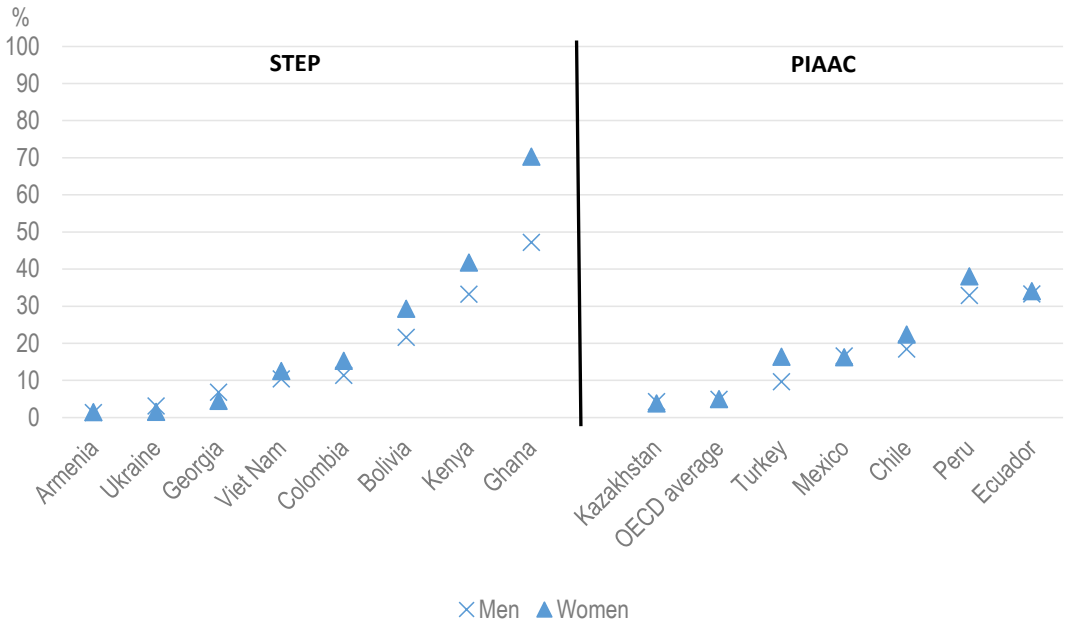
Table 9. Average literacy proficiency, by gender

	Men		Women	
	Mean score	S.E.	Mean score	S.E.
STEP				
Armenia	255.5	(2.1)	256.2	(1.7)
Bolivia	208.2	(3.8)	192.8	(3.2)
Colombia	241.9	(3.2)	230.2	(3.9)
Georgia	241.5	(2.4)	247.9	(1.7)
Ghana	162.1	(4.0)	120.3	(3.4)
Kenya	180.5	(3.1)	164.4	(3.2)
Ukraine	266.0	(3.5)	269.8	(1.9)
Viet Nam	246.8	(2.5)	239.8	(2.4)
PIAAC				
Chile	223.9	(2.5)	216.4	(2.8)
Ecuador	197.0	(1.3)	195.8	(1.3)
Kazakhstan	248.0	(1.4)	250.2	(1.1)
Mexico	222.3	(1.4)	220.9	(1.1)
Peru	198.6	(1.2)	192.6	(1.1)
Turkey	232.0	(1.6)	220.9	(1.3)
OECD average	**267.0**	**(0.2)**	**265.3**	**(0.2)**

Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

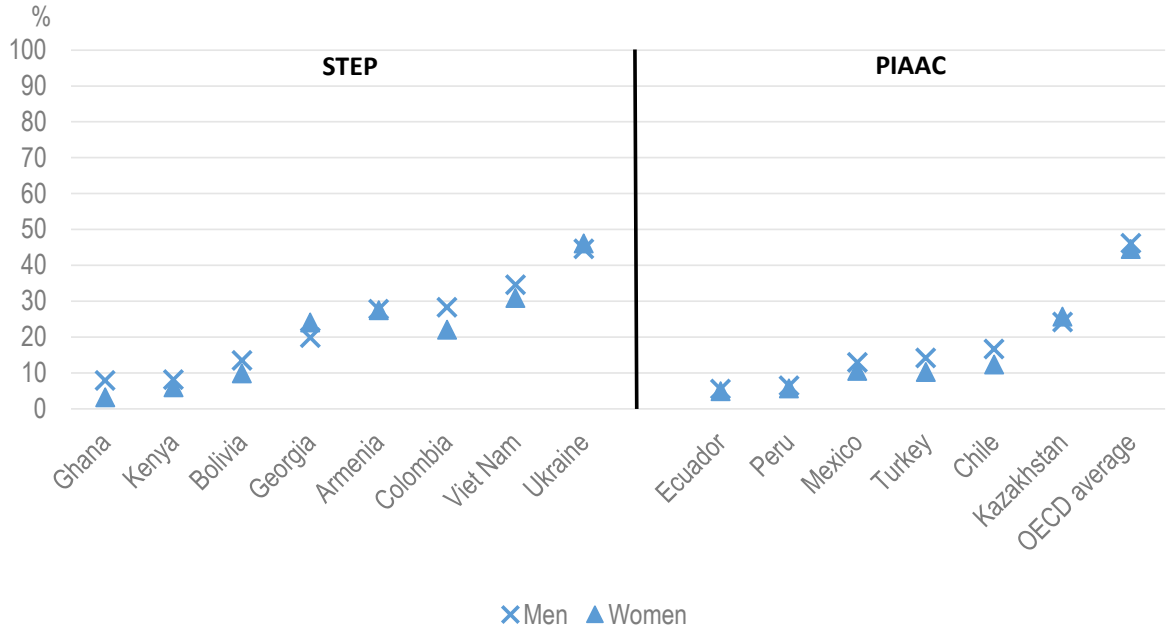
Figure 24 and Figure 25 show that gender gaps are rather similar along the proficiency distribution, with the exception of Bolivia, Ghana and Kenya, where women are much more likely than men to score below Level 1 (while having similar probability of scoring at Level 3 or above).

Figure 24. Percentage of adults scoring below Level 1, by gender



Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Figure 25. Percentage of adults scoring at Level 3 or above, by gender



Notes: The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

4.5. Conclusions

As it has been noted, comparisons between results of STEP and PIAAC, as well as between the countries participating in STEP need to take into account, among other things, the fact that different target populations were covered in the two studies. Focussing on the relationships between literacy proficiency and background characteristics, the basic patterns observed in the analysis of multiple rounds of PIAAC data are confirmed in STEP. Educational attainment is strongly correlated with literacy proficiency, and in all countries and economies, more educated adults display higher scores in the literacy assessment. Proficiency tends to decrease with age, although the shape of the age-proficiency profile is much more country-specific, because cross-sectional surveys like PIAAC and STEP do not allow to separately identify age and cohort effects. Each country has its own history and timing of educational expansion, resulting in differences in the average educational attainment of adults born at different points in time. In lower- and middle-income countries, the expansion of educational opportunities tends to be a more recent phenomenon: as a result, younger cohorts tend to be much more educated than older cohorts, and skills differences between younger and older respondents tend to be larger than in more developed western countries (where young and old adults have more similar levels of education). Finally, gender differences in proficiency are small everywhere, although they do tend to be slightly larger in low- and middle-income countries (which are over-represented in STEP).

5. Literacy proficiency and labour market outcomes

An important motivation for conducting assessments of adults' skills is the belief that information-processing skills are necessary for adults to thrive in modern societies and modern labour markets. Analysis of PIAAC data has indeed convincingly shown that literacy, numeracy and problem-solving skills have high returns in modern labour markets, meaning that more skilled individuals tend to earn higher wages and face lower risks of unemployment (Hanushek et al., 2015[28]; OECD, 2019[29]). Moreover, skills have been shown to be associated also with important non-economic outcomes, such as trust, political efficacy and self-reported health (Kakarmath et al., 2018[30]; Borgonovi and Burns, 2015[31]; Borgonovi and Pokropek, 2017[32]).

This chapter will present some evidence on the correlation between literacy proficiency and economic outcomes in PIAAC and STEP countries. While Chapter 4 was mainly focused on the supply of skills (i.e. how skills are built through education, and how they evolve with age), the focus of this chapter will be on the demand for skills, i.e. to what extent the information-processing skills are demanded in the labour market, and how much adults are rewarded for their proficiency.

As in Chapter 4, the analysis will focus on the comparison of relationship between literacy proficiency and adult outcomes. Finding coherent results across the two studies would provide some evidence for the validity of the PIAAC instrument in assessing a similar construct across a wide range of contexts. Results will be presented for all STEP countries and for a number of selected PIAAC countries (focussing in particular on upper-middle-income countries, which are more similar to STEP countries in terms of level of economic development and estimated adults' proficiency).

5.1. The demand to use literacy skills

The demand for skills is normally linked to the sectoral and occupational structure of the economy, and on its level of development. Economic development is normally driven by technological change, and skills are required not only to advance the technological frontier, but often also to use new technologies. High economic returns to cognitive skills are an important incentive to acquire education and increase proficiency; on the other hand, an insufficiently skilled workforce can be a constraint to economic development and to the creation of jobs that can make good use of information-processing skills. The demand for literacy and information processing skills more generally in employment and in social life more broadly represents an important incentive for building and maintaining proficiency over the life course.

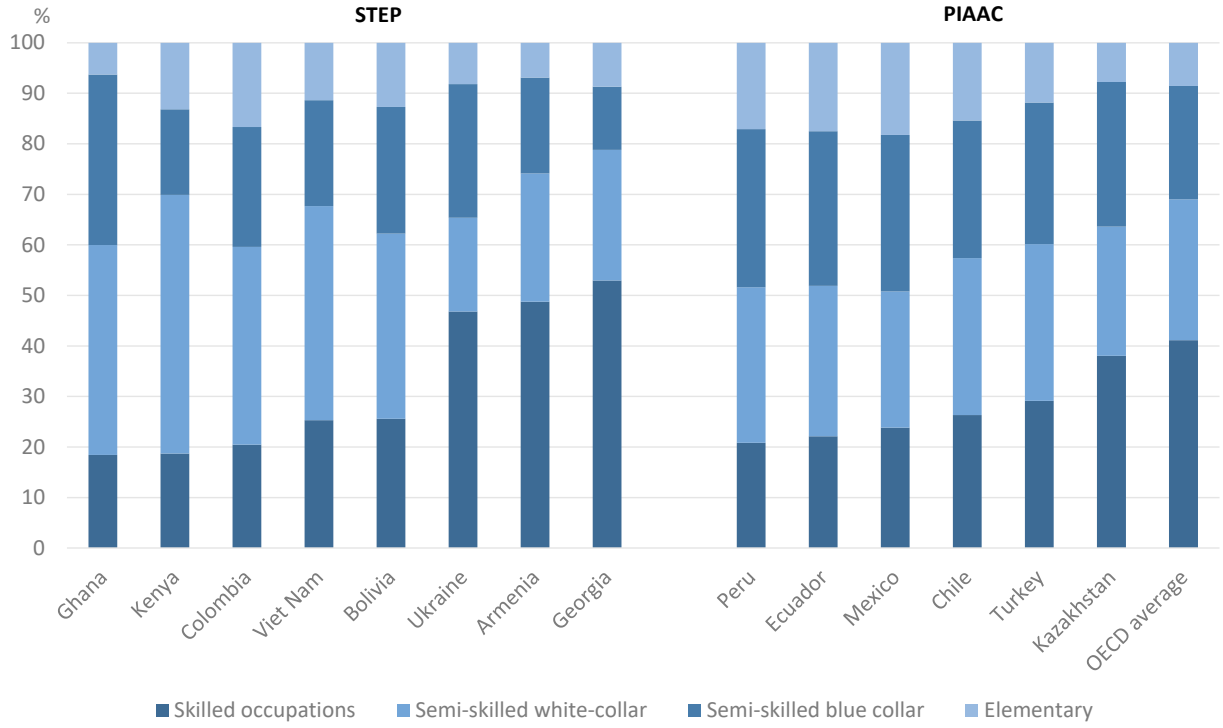
Chapter 1 has shown that PIAAC and STEP countries are at different stages of economic development, and this is likely to result in differences in the economic structure and in the demand for skills (although the urban areas that are surveyed in STEP countries are more likely to be more economically advanced and thus more similar to the OECD countries and economies surveyed in PIAAC).

One way to assess these types of structural differences is to examine the share of adults employed in different occupations and rely on an ex-ante classification of occupations on the basis of their skills content. This is unfortunately another example where the different sampling frame adopted in STEP prevents a direct comparison between the two studies, as some of the differences in the distribution of adults across occupations will partly be due to the fact that, in any country, the mix of occupations present in urban areas is different from the mix of occupations present in rural areas. On the other hand, this would be less of a problem when analysing the link between proficiency and occupational outcomes.

The International Labour Organization (ILO) classifies occupations on the basis of their required skill level and skill specialisation: skill level is "*a function of the complexity and range of tasks and duties to be performed in an occupation*", while skill specialisation is considered in terms of the field of knowledge required, the tools and machinery used, the materials worked on or with, and the kinds of goods and services produced (ILO, 2012[33]). At the highest level, occupations are classified in four groups: skilled occupations, semi-skilled white-collar occupations; semi-skilled blue collar occupations; and elementary occupations. Skilled occupations include, for instance, managers and professionals; examples of semi-skilled white-collars jobs are clerks and service and sale workers; semi-skilled blue collars include craft workers and skilled agricultural workers; elementary occupations include, among others, cleaners, street vendors, and labourers in agriculture, transport or construction. As both PIAAC and STEP code respondents' occupation according to the International Standard Classification of Occupations (ISCO) classification, this information is comparable across countries and economies that participated in the two studies.

Figure 26 shows the percentage of employed adults in different occupations. With the exception of Ukraine, Armenia, and Georgia, adults surveyed in other STEP countries and in PIAAC countries and economies with lower levels of GDP per capita and literacy proficiency are less likely to work in skilled occupations than adults living in the average OECD country that participated in PIAAC.

Figure 26. Share of workers in different occupations

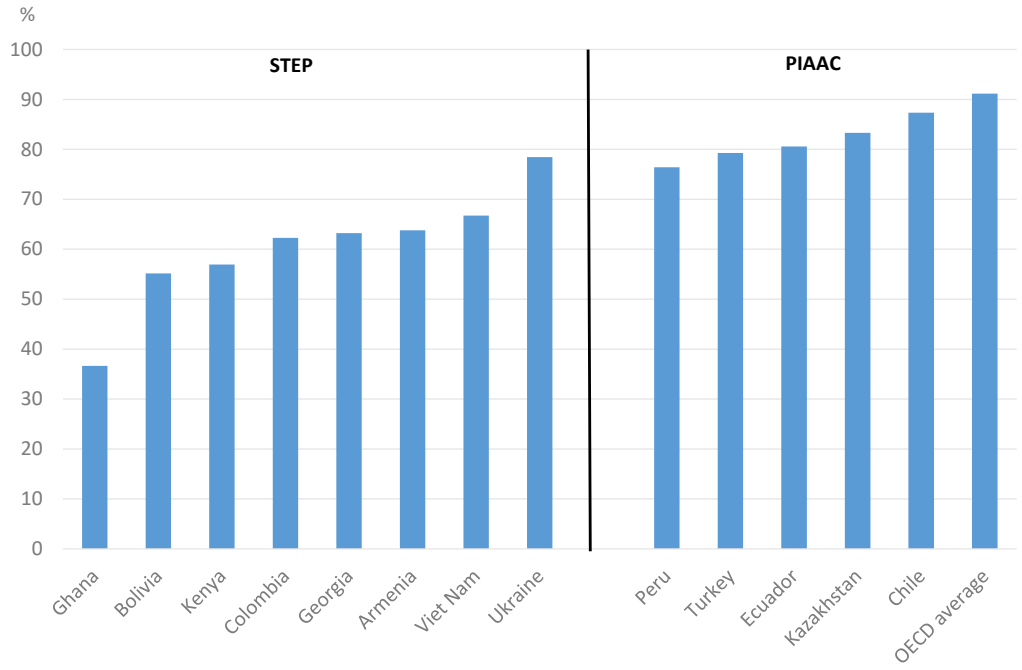


Notes: Percentage of employed adults in different types of occupations, classified according to (ILO, 2012[33]). The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

One shortcoming of using the ILO classification of occupations is that an implicit assumption is made that all workers in an occupation perform similar tasks in all countries and economies. There is instead ample evidence of large within-occupation heterogeneity in tasks carried out at work, which is arguably the result of differences in the way work is organised in different firms and countries (Autor and Handel, 2013[34]). This is one reason the Background Questionnaire for PIAAC and STEP included a range of questions on the tasks performed at work (from which the use of different skills can be inferred). As explained in Chapter 2, differences in the way such questions are asked prevent to fully exploit the range of information in a comparable way across the two surveys. However, some basic conclusions can still be drawn, for example, regarding the use of reading (in a generic sense) and the use of computer at work. Even as far as reading is concerned (without taking into account the complexity of the texts workers have to engage with), a significant difference emerges between many STEP countries and the average OECD country that participated in PIAAC: while more than 9 workers out of 10 engage with written texts on average across OECD countries and economies, less than 4 do so in Ghana (Figure 27). The shares are higher in other STEP countries, but they are normally below 70%, with the only exception of Ukraine.

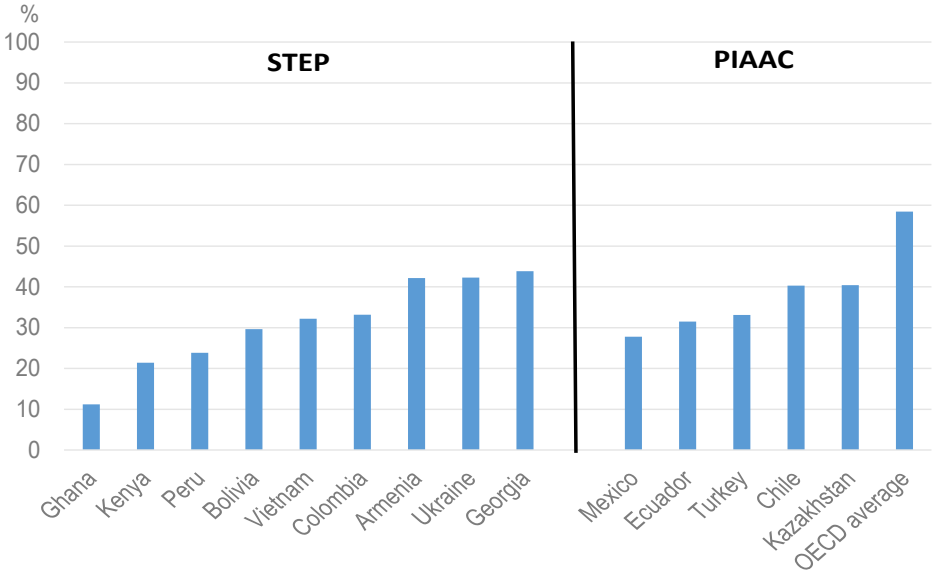
Similarly, large differences are observed in the share of workers who use a computer at work: almost 60% on average across OECD countries, as compared to about 10% in Ghana, around 30% in Bolivia, Viet Nam and Colombia, and around 40% in Armenia, Ukraine and Georgia (Figure 28).

Figure 27. Share of adults reading at work



Notes: Percentage of employed adults who report to be reading at work. The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Figure 28. Share of adults using a computer at work



Notes: Percentage of employed adults who report to be using a computer at work. The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

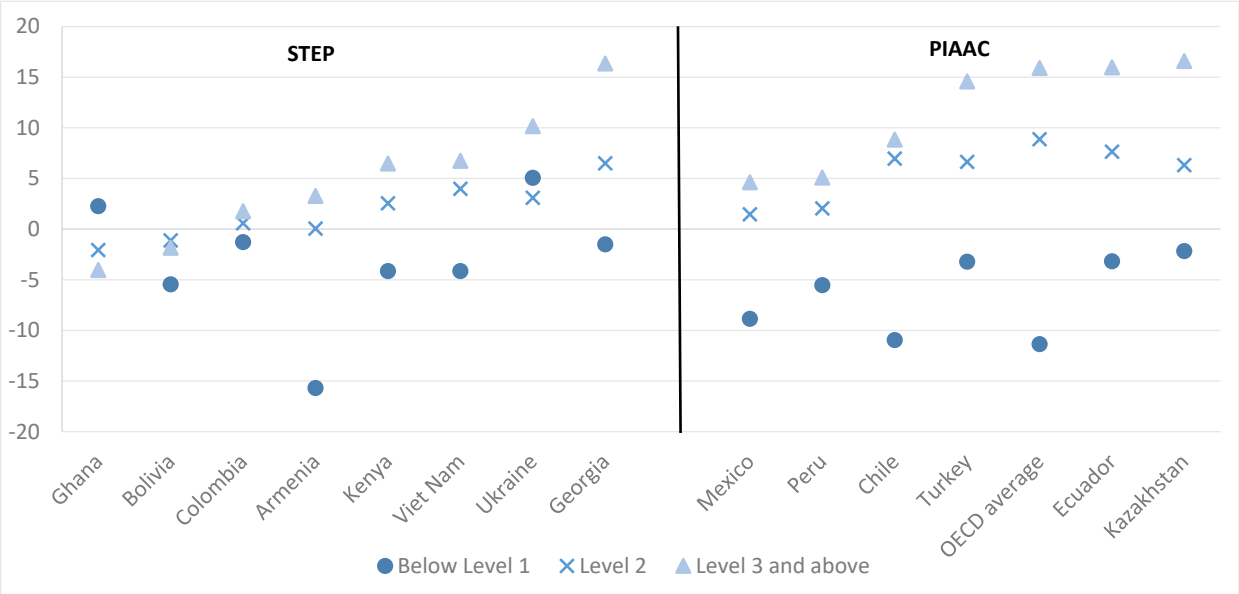
5.2. Literacy proficiency and labour market outcomes

Even in economies where the share of jobs demanding workers to make use of information processing skills is relatively low, acquiring those skills can still be a profitable investment from a purely economic perspective of individual workers, as long as demand for skills exceeds supply. From a macroeconomic perspective, the argument is stronger, to the extent that human capital (skills and education) generate benefits for the entire society that go beyond the individual economic returns (Winters, 2018[35]).

Figure 29 shows that in most countries prime-aged adults (aged between 25 and 54 years, who have the highest attachment to the labour market and are mostly no longer in education and not yet retired) are more likely to be employed if they have higher proficiency in literacy.

Figure 29. Literacy proficiency and the probability of being employed

Percentage points differences in the probability of being employed compared with adults at Level 1



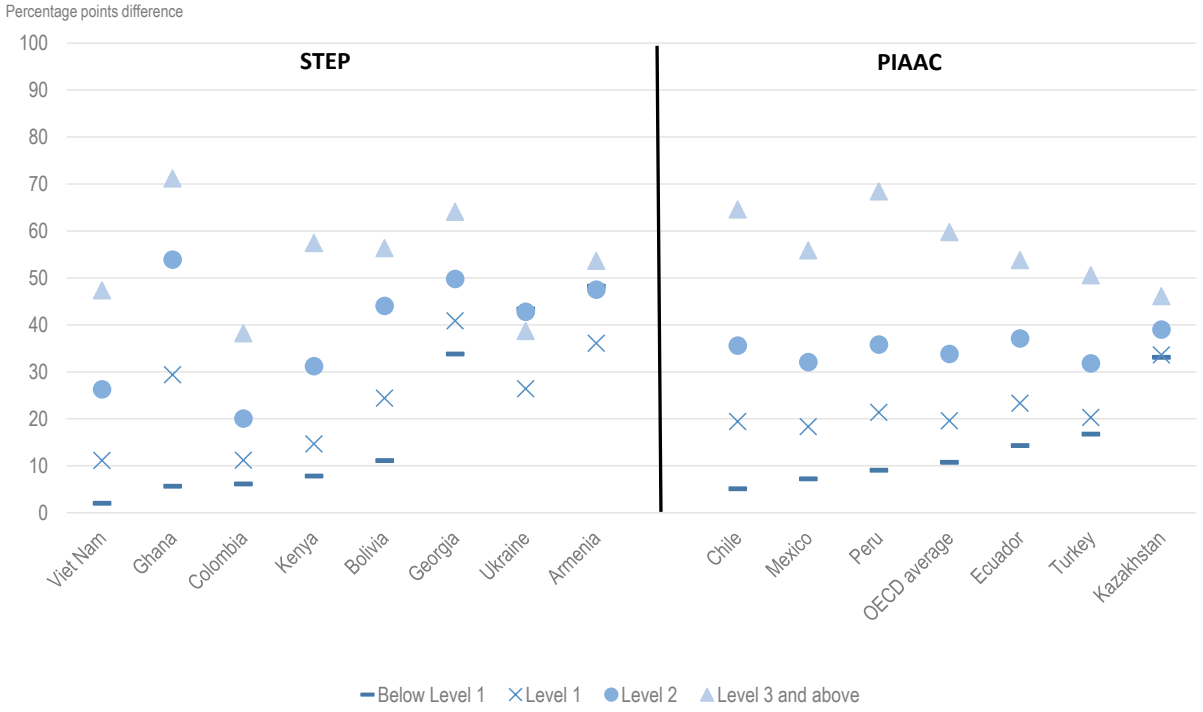
Notes: The sample is restricted to adults aged 25-54, and the differences are estimated controlling for age and gender. The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

This is not true everywhere, though. In Ghana, adults scoring at Level 3 or above are actually less likely to be employed than adults scoring at lower levels, which suggests a scarcity of high-skills jobs. In Bolivia and Colombia there are also very small differences in employment rates between adults at different levels of literacy proficiency.

Conditional on being in employment, proficiency in literacy is associated with the type of jobs adults have access to. It would be reasonable to expect that more skilled adults tend to work in occupations in which their skills are in higher demand. Figure 30 shows that adults with higher levels of proficiency are normally much more likely to work in skilled occupations than lower-skilled adults. Similar results hold for elementary and middle-skills jobs, where lower-skilled adults are more likely to be employed. Exceptions to this pattern are Armenia, Ukraine and Kazakhstan, where the chances for employed adults to work in a skilled occupation appear to be much less related to their proficiency level.

Figure 30. Employed adults in skilled occupations, by proficiency level

Adults aged 25-54



Notes: The sample is restricted to adults aged 25-54. The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

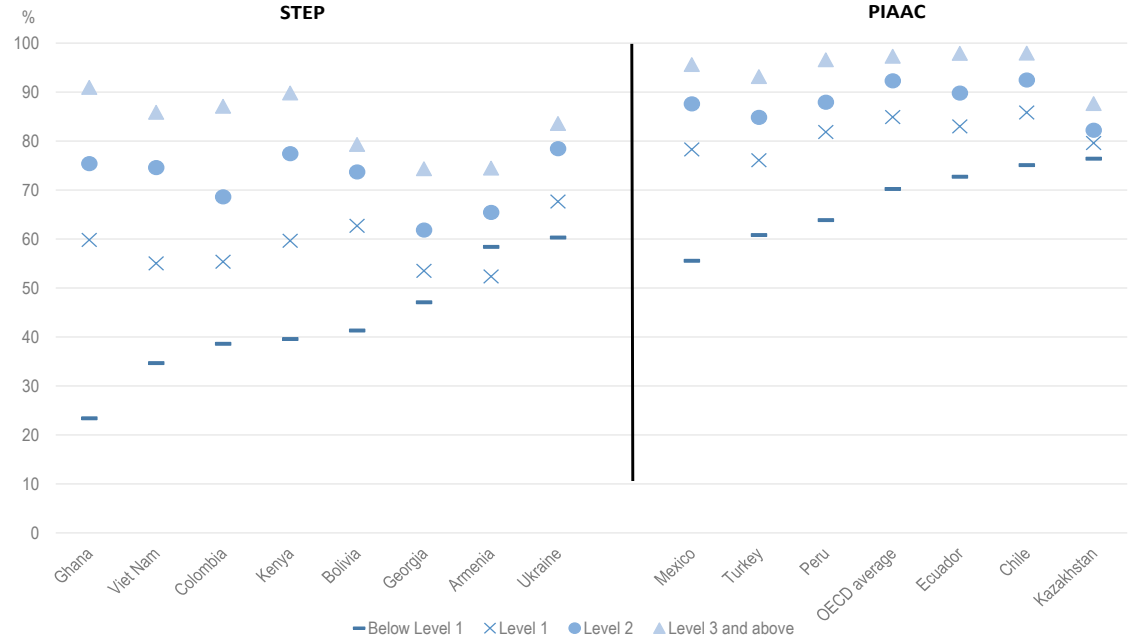
The positive assortative matching of skilled workers with skilled jobs is reflected not only in abstract occupational classifications, but also in what workers report they are actually doing on the job. Figure 31 shows that employed adults scoring at higher levels of proficiency are more likely to be engaging with written texts at work, although this relationship is less strong in Armenia, Kazakhstan and Ukraine.

Similarly, more proficient adults are more likely to be using a computer at work, as shown in Figure 32. Armenia, Kazakhstan and Ukraine are again partial exceptions to this pattern.

Ultimately, these differences in occupations are reflected in large differences in wages, as reported in Figure 33. Returns to higher levels of proficiency are particularly high in countries with low average scores, like Colombia, Ghana and Kenya. Consistent with the results on occupational outcomes, they are much smaller in Armenia, Georgia and Ukraine.

Figure 31. Employed adults reading at work, by proficiency level

Adults aged 25-54

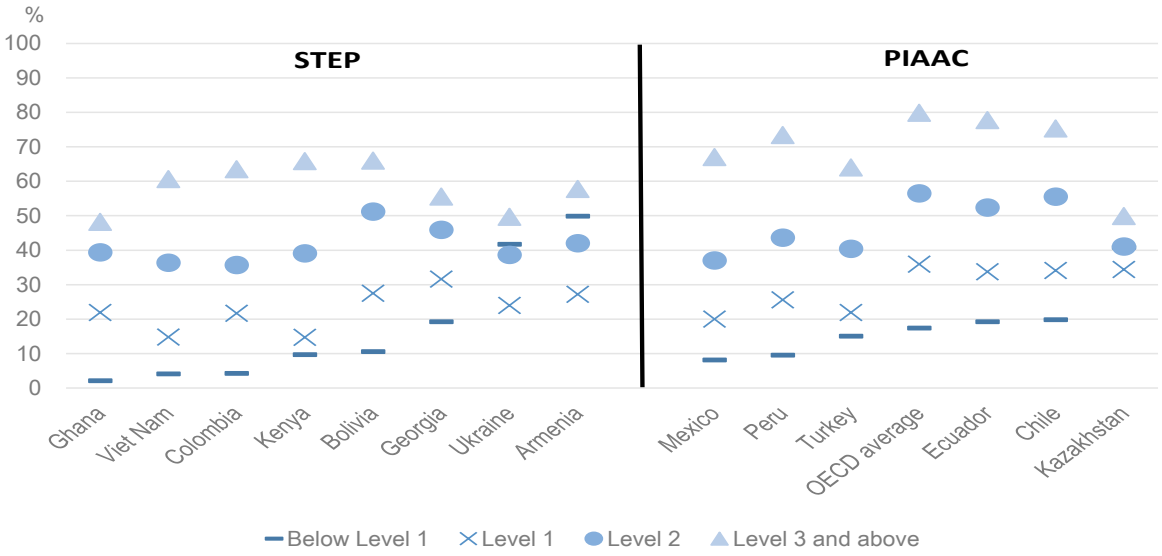


Notes: The sample is restricted to adults aged 25-54. The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.

Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Figure 32. Employed adults using computer at work, by proficiency level

Adults aged 25-54

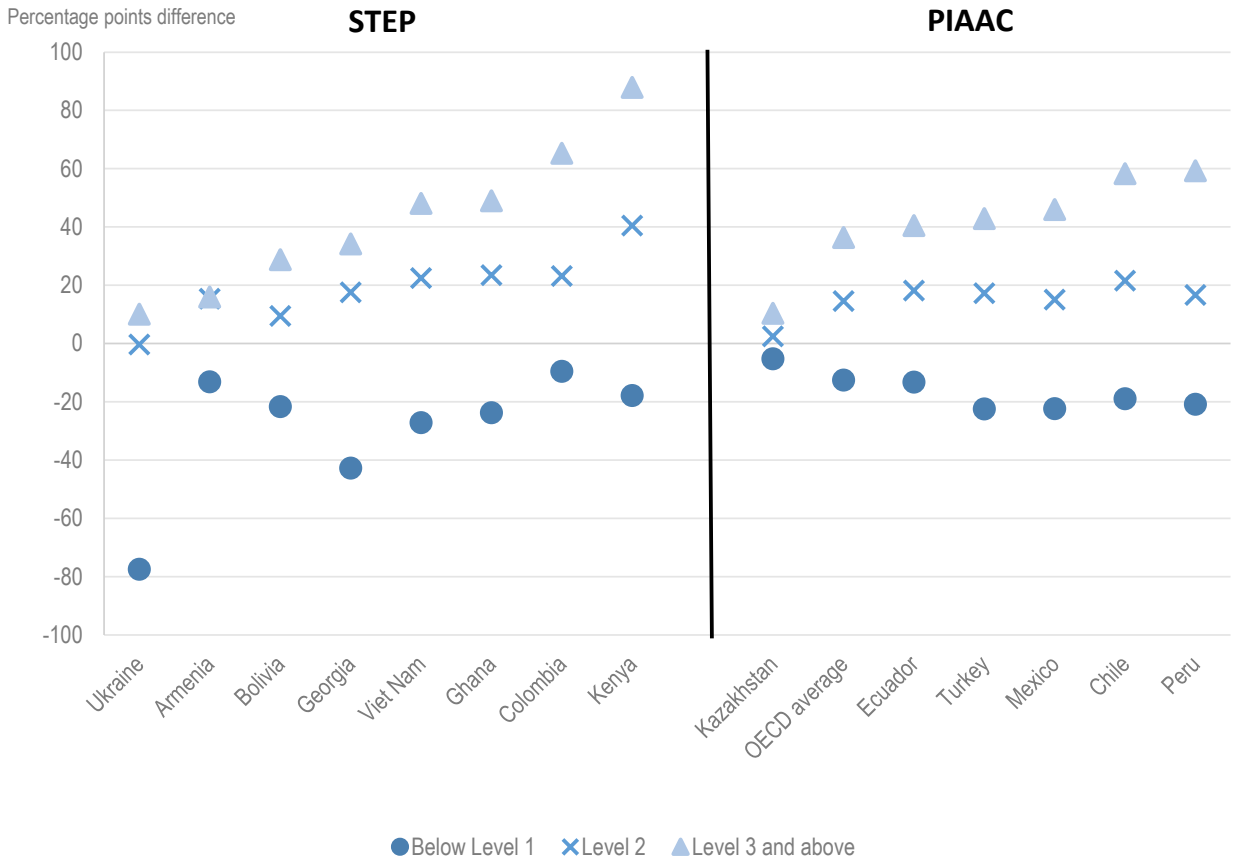


Notes: The sample is restricted to adults aged 25-54. The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.

Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

Figure 33. Wage returns to literacy

Employed adults aged 25-54



Notes: The figure reports the estimated percentage differences in hourly wages with respect to adults scoring at Level 1. The regression controls for age and gender and the sample is restricted to adults aged 25-54. The data for STEP countries only refer to the adult population living in urban areas. The OECD average refers to the average score of all OECD member countries and economies that participated in PIAAC.
Sources: OECD ([24]) Survey of Adult Skills (PIAAC), http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020) and World Bank ([15]), STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020).

5.3. Conclusions

Countries and economies that participated in PIAAC and STEP vary substantially in their level of economic development and in their economic and occupational structure. This also means that the demands on adults in terms of literacy proficiency are also quite different, which is an important aspect to take into account when interpreting the results obtained by participants in the literacy assessment.

The relationship between skills and labour market outcomes remains however strong in most countries and economies, consistent with well-known findings from PIAAC. Adults scoring at higher levels in the literacy assessment are more likely to be employed, to work in more skilled occupations, to make use of a computer and to engage with written texts at work. Ultimately, this is reflected in higher wages earned by more skilled adults.

6. Conclusions

This report has described similarities and differences between PIAAC and STEP and has presented some analysis of data from the two studies.

A first conclusion that can be drawn from the report is that, in spite of the many similarities, the differences that exist between the two surveys are significant, and prevent a joint analysis of results from the two studies. These differences relate primarily to the way the studies were implemented, and secondarily to the instruments they use.

In terms of implementation, the sampling strategy used in STEP strongly prevents any type of international comparison, not only between STEP and PIAAC countries and economies, but also between countries that participated in STEP. As the definition of "urban areas" from which units are sampled is national, the results will necessarily be influenced by the characteristics of said urban areas in each particular country. This is just an example of the more general point that achieving cross-country and cross-cultural comparability requires a careful analysis of all stages of a survey, and that using a common set of instruments is typically not enough.

In terms of instrumentation, the main differences relate to the content of the Background Questionnaire. While it is certainly desirable to customise to the same extent the content of the Background Questionnaire to different national contexts, it would also be advisable to uniform the way the same construct is elicited in different surveys, especially when they use the same instruments to assess adults' skills. This is particularly evident in the different ways PIAAC and STEP use to elicit information on the use of skills at work and in everyday life.

A second, more positive, conclusion can be drawn concerning the feasibility of assessing adult literacy on a global scale. There is no evidence that the PIAAC instruments do not function or function differently in low- and middle-income countries that participated in STEP. This can be seen in the analysis of response curves for common items administered in the PIAAC and STEP core, as well as in the fact that the estimated proficiency correlates in similar ways with background characteristics and is related in similar ways with labour market outcomes. The issues identified in this report have to do with the insufficient coverage of the assessment at the lower end of the skills distribution, which results in an excessive reliance on psychometric models to estimate scores in the presence of a very little amount of information on what respondents are actually able to do.

There is therefore considerable scope to improve the measurement of proficiency for adults at the bottom end of the skills distribution, and this is the area where efforts should concentrate more urgently. While the PIAAC instruments (the assessment and the theoretical framework) can indeed form a base for benchmarking adults' skills on a global scale, they should however be expanded to provide better coverage and higher resolution of the bottom end of the skills distribution. This endeavour can certainly be facilitated by technological developments. The increasingly widespread use and accessibilities of digital devices will open up new possibilities to implement more efficient and more sophisticated assessments, which will increasingly rely on adaptive designs.

References

Autor, D. and M. Handel (2013), "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages", *Journal of Labor Economics*, Vol. 31/S1, pp. S59-S96, http://dx.doi.org/10.1086/669332. [34]

Barro, R. and J. Lee (2013), "A new data set of educational attainment in the world, 1950-2010", *Journal of Development Economics*, Vol. 104, pp. 184-198, http://dx.doi.org/10.1016/j.jdeveco.2012.10.001. [4]

Borgonovi, F. and T. Burns (2015), "The educational roots of trust", *OECD Education Working Papers*, No. 119, OECD Publishing, Paris, https://dx.doi.org/10.1787/5js1kv85dfvd-en. [31]

Borgonovi, F. and A. Pokropek (2017), "Mind that gap: The mediating role of intelligence and individuals' socio-economic status in explaining disparities in external political efficacy in 28 countries", *Intelligence*, Vol. 62, pp. 125-137, http://dx.doi.org/10.1016/j.intell.2017.03.006. [32]

Gaëlle, P. et al. (2014), "STEP Skills Measurement Surveys. Innovative Tools for Assessing Skills", *Social Protection & Labor Discussion Paper*, No. 1421, World Bank. [13]

Handel, M. (2016), "What do people do at work?", *Journal for Labour Market Research*, Vol. 49/2, pp. 177-197, http://dx.doi.org/10.1007/s12651-016-0213-1. [17]

Handel, M. (2008), "Measuring Job Content: Skills, Technology, and Management Practices", *Institute for Research on Poverty Discussion Paper*, No. 1357-08, University of Wisconsin, Madison, WI. [18]

Hanushek, E. et al. (2015), "Returns to skills around the world: Evidence from PIAAC", *European Economic Review*, Vol. 73, pp. 103-130, http://dx.doi.org/10.1016/j.euroecorev.2014.10.006. [28]

ILO (2012), *International Standard Classification of Occupations: ISCO-08*, International Labour Office, Geneva, http://www.ilo.org/public/english/bureau/stat/isco/index.htm (accessed on 8 February 2018). [33]

Kakarmath, S. et al. (2018), "Association between literacy and self-rated poor health in 33 high-and upper-middle-income countries", *OECD Education Working Papers*, No. 165, OECD Publishing, Paris, https://dx.doi.org/10.1787/7aaeac27-en. [30]

Lyberg, L., K. Cibelli Hibben and B. Pennell (2018), "Applying the total survey error framework to PIAAC", *Quality Assurance in Education*, Vol. 26/2, pp. 153-168, http://dx.doi.org/10.1108/QAE-07-2017-0035. [7]

OECD (2019), *Skills Matter: Additional Results from the Survey of Adult Skills*, OECD Skills Studies, OECD Publishing, Paris, https://dx.doi.org/10.1787/1f029d8f-en. [27]

OECD (2019), "The outcomes of investments in skills", in *Skills Matter: Additional Results from the Survey of Adult Skills*, OECD Publishing, Paris, https://dx.doi.org/10.1787/19de3145-en. [29]

OECD (2019), *The Survey of Adult Skills : Reader's Companion, Third Edition*, OECD Skills Studies, OECD Publishing, Paris, https://dx.doi.org/10.1787/f70238c7-en. [14]

OECD (2016), *Technical Report of the Survey of Adult Skills (PIAAC) (2nd Edition)*, OECD Publishing, Paris, https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf (accessed on 22 May 2019). [23]

OECD (2016), *The Survey of Adult Skills: Reader's Companion, Second Edition*, OECD Skills Studies, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264258075-en. [22]

OECD (2013), *Technical Report of the Survey of Adult Skills (PIAAC)*, OECD Publishing, Paris, https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf. [9]

OECD (2013), "The Methodology of the Survey of Adult Skills (PIAAC) and the Quality of Data 3", in *The Survey of Adult Skills: Reader's Companion*, OECD Publishing, Paris, https://doi.org/10.1787/9789264204027-6-en (accessed on 1 April 2020). [6]

OECD (n.d.), *Survey of Adult Skills (PIAAC)*, http://www.oecd.org/skills/piaac/data/ (accessed on 13 September 2020). [24]

PIAAC Literacy Expert Group (2009), "PIAAC Literacy: A Conceptual Framework", *OECD Education Working Papers*, No. 34, OECD Publishing, Paris, https://dx.doi.org/10.1787/220348414075. [19]

Puga, D. (2010), "The Magnitude and Causes of Agglomeration Economies", *Journal of Regional Science*, Vol. 50/1, pp. 203-219, http://dx.doi.org/10.1111/j.1467-9787.2009.00657.x. [10]

Richmond, M., C. Robinson and Sachs-Israel Margarete (eds.) (2008), *The Global literacy challenge: a profile of youth and adult literacy at the mid-point of the United Nations Decade 2003-2012 - UNESCO Bibliothèque Numérique*, UNESCO , Paris, https://unesdoc.unesco.org/ark:/48223/pf0000163170 (accessed on 17 April 2020). [25]

Roseth, V., A. Valerio and M. Gutiérrez (2016), *Education, Skills, and Labor Market Outcomes: Results from large-scale adult skills surveys in urban areas in 12 countries*, World Bank, Washington, D.C., http://hdl.handle.net/10986/24276 (accessed on 6 July 2020). [1]

Survey Research Center (2016), *Guidelines for Best Practice in Cross-Cultural Surveys Full Guidelines*, Survey Research Center, Institute for Social Research, University of Michigan, http://www.ccsg.isr.umich.edu/ (accessed on 1 April 2020). [5]

UNESCO (n.d.), *World Inequality Database on Education*, https://unesco-wide-production.herokuapp.com/ (accessed on 1 April 2020). [12]

UNESCO Institue for Statistics (n.d.), *UIS.stat*, http://data.uis.unesco.org/ (accessed on 16 April 2020). [26]

UNESCO, I. (2018), *Data to Nurture Learning SDG 4 Data Digest 2018*, UIS, Montreal. [21]

Von Davier, M., E. Gonzalez and R. Mislevy (2009), "What are plausible values and why are they useful?", in *Issues and Methodologies in Large-Scale Assessments*. [20]

Winters, J. (2018), "Do Higher Levels of Education And Skills in an Area Benefit Wider Society?", *World of Labor*, No. 130, IZA, http://dx.doi.org/10.15185/izawol.130.v2. [35]

World Bank (2019), *New country classifications by income level: 2019-2020*, https://blogs.worldbank.org/opendata/new-country-classifications-income-level-2019-2020 (accessed on 6 May 2020). [2]

World Bank (2014), "STEP skills measurement surveys : innovative tools for assessing skills", *Social protection and labor discussion paper*, No. 1421, World Bank Group, Washington, DC, http://documents.worldbank.org/curated/en/516741468178736065/STEP-skills-measurement-surveys-innovative-tools-for-assessing-skills (accessed on 1 April 2020). [8]

World Bank (n.d.), *Health, Nutrition and Population Statistics Database*, https://databank.worldbank.org/source/health-nutrition-and-population-statistics (accessed on 1 April 2020). [16]

World Bank (n.d.), *STEP Skills Measurement Program*, https://microdata.worldbank.org/index.php/catalog/step (accessed on 1 April 2020). [15]

World Bank (n.d.), *World Development Indicators*, https://databank.worldbank.org/source/world-development-indicators (accessed on 28 October 2019). [3]

Young, A. (2013), "Inequality, the Urban-Rural Gap, and Migration", *The Quarterly Journal of Economics*, Vol. 128/4, pp. 1727-1785, http://dx.doi.org/10.1093/qje/qjt025. [11]