



BY WAY OF CONCLUSION

OECD seminar on managing quality in large scale
assessments, 11-12 May, 2017, Paris
William Thorn, OECD



-
- 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.



Structure of the presentation

- Context: why was this seminar organised?
- Relevant information in PIAAC
- Some messages from the seminar
- Implications for PIAAC/further work

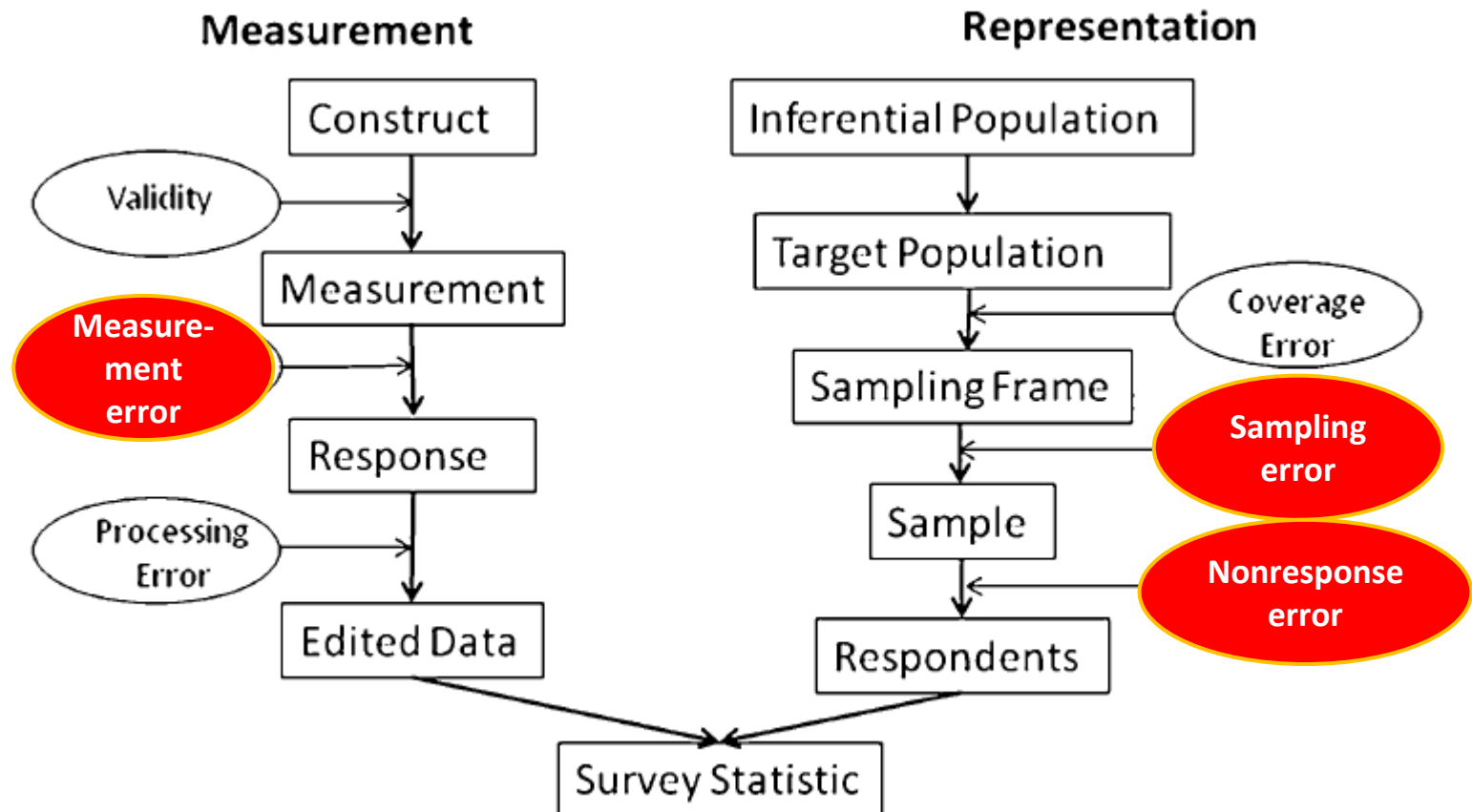


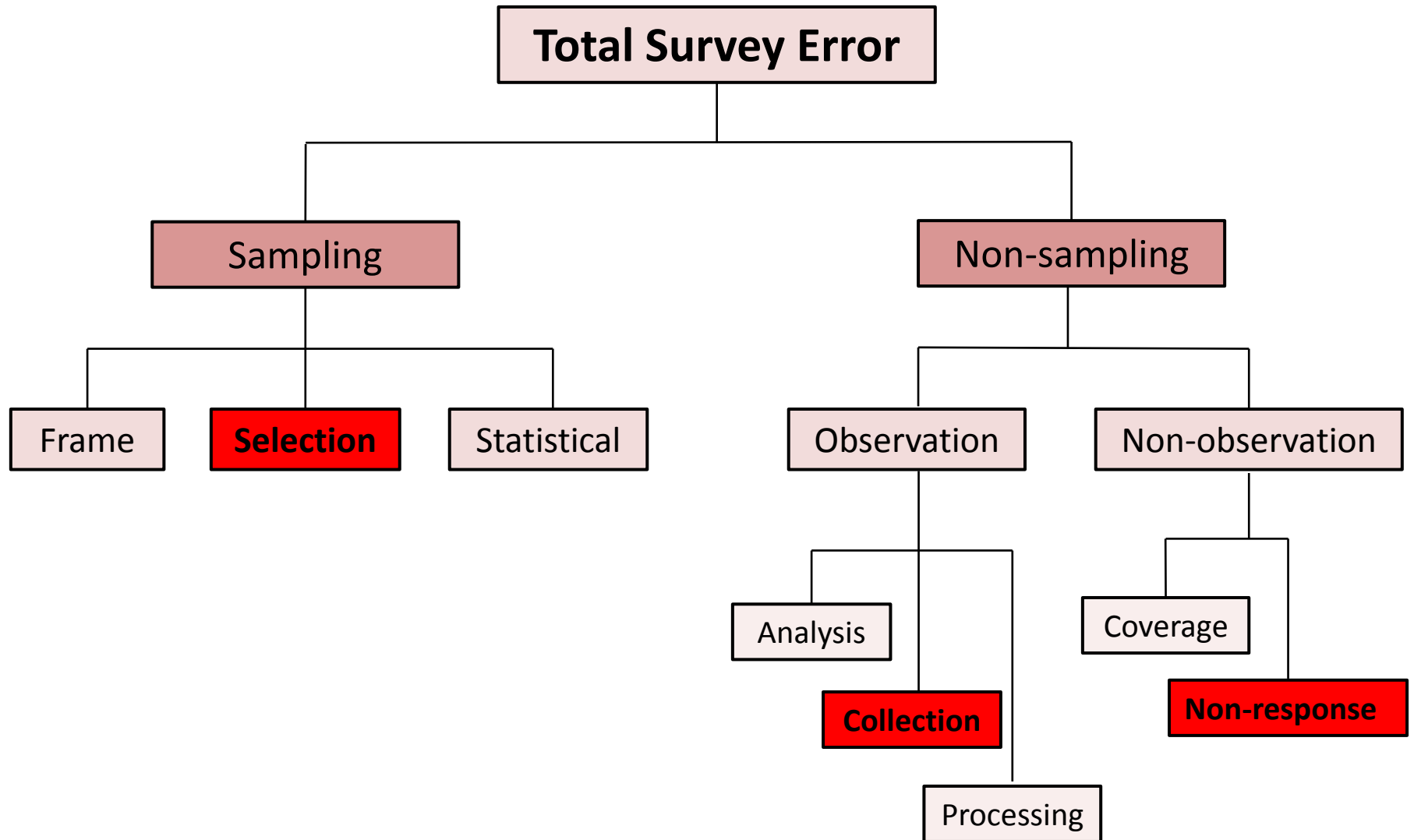
Context

- Interest in better understanding and managing sources of error linked to field operations and data collection
- Developments in IT
 - Lots more information (paradata) to exploit with CAPI/CBA delivery
 - How to make best use of capacity of CAPI/CBA in QC and QA and detection and remediation of data problems



Context: sources of error







Measurement	Representation
Validity Frameworks (assessment, BQ) Instruments	Coverage error Sample frame
Measurement error Translation Data collection Scaling	Sampling error Sample design Sample selection (households and/or individuals)
Processing error Coding Scoring Data processing	Non-response error Persons Items



Low risks for quality: Measurement

- Validity
 - Assessment frameworks (defines construct and features of the measures)
 - Item development linked to framework
- Measurement errors
 - Translation undertaken under strict guidelines with centralised verification
- Processing errors
 - Data entry minimised (CAPI, CBA)
 - Automatic scoring (CBA)
 - Manual scoring (PBA only) reliability assessed
 - Well developed protocols for data transfer, data cleaning, editing



Low risks to quality: Representation

- Coverage error
 - Most countries in PIAAC have sample frames of reasonable quality
 - Maximum of 5% exclusions
- Sampling error
 - Guidance from consortium in developing sampling plans
 - Design effects minimised



High risks to quality: Measurement

- Collection errors
 - Interviewer (failure to follow protocols, influence on respondents, satisficing, falsification)
 - Setting (distractions, lack of confidentiality)
 - Respondent (cognitive, presentational, motivational)
 - Data collection organisation (training, management of operations, falsification)



High risks to quality: Representation

- Sampling errors
 - Selection of households and/or respondents by the interviewer
- Non-response
 - Non-response rate (persons) is high in many countries
 - Unavailable/uncontactable respondents
 - Other (e.g. language, literacy, disability)



Main areas of concern about data quality in PIAAC

- Non-response rates
- Literacy-related non-response
- Data irregularities (two countries)
- Differences in operational procedures in the field
- Large Design effects (two countries)



Context: Developments in ICT

- PIAAC was the first international large-scale assessment to be delivered predominantly in CAPI/CBA mode
- CAPI/CBA enormously increased the amount of information available about data collection and the behaviour of test-takers.
 - Timing and other process data.
 - Only just starting to explore these data
- CAPI/CBA has increased the opportunities to detect and act to resolve emerging problems in a timely fashion and, possibly, to prevent them.

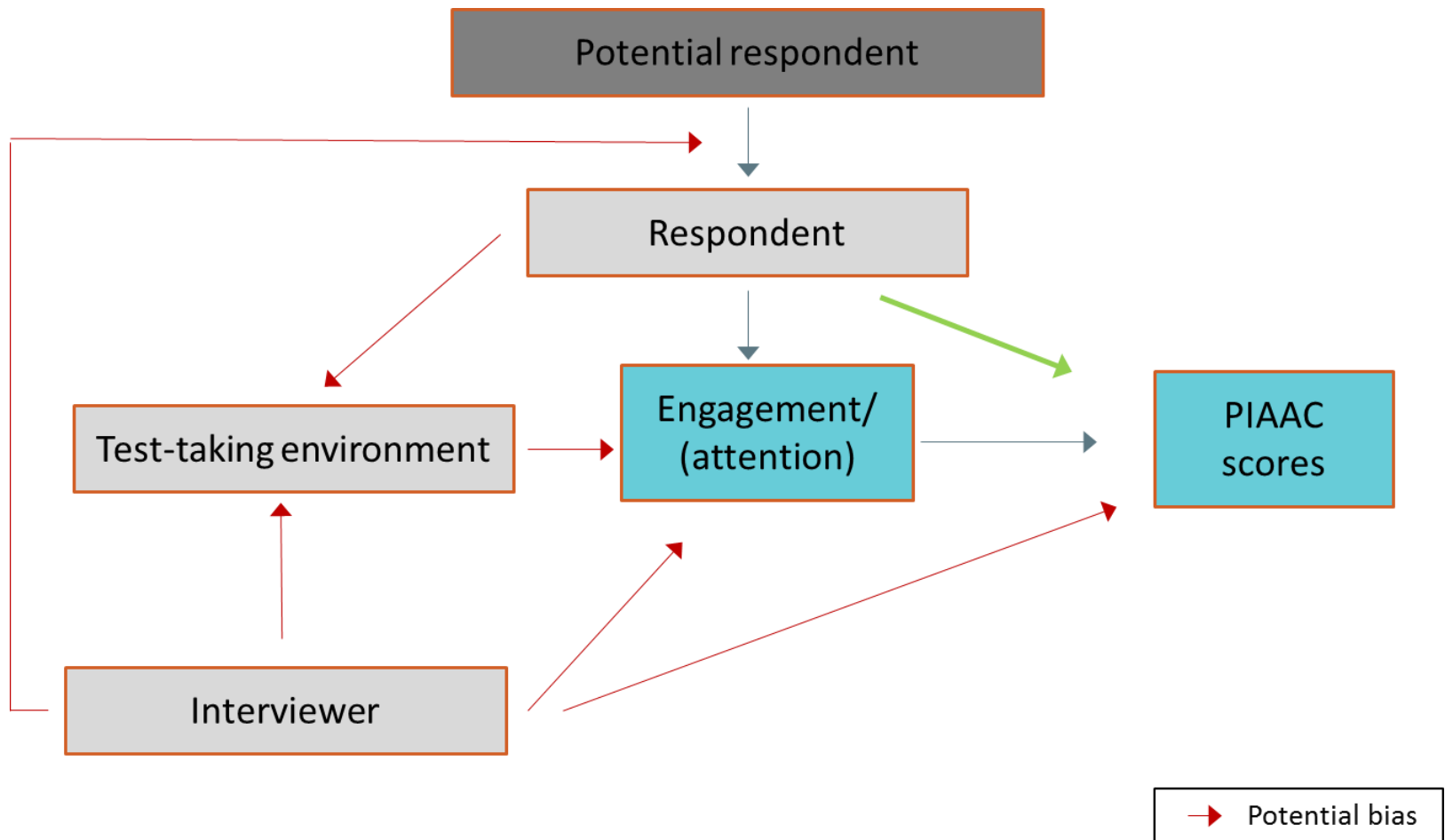


Information in PIAAC (Cycle 1)

- Interviewer IDs linked to respondents
 - Assignment numbers
- Logs of interactions with computer (BQ and cognitive assessment)
 - Time stamps plus other info
- Observation module (ZZ questions)
 - Presence of other person during interview
 - Assistance from an other person
 - Respondent asked for help
 - Respondents complained about the duration of the assessment
 - Room in which assessment took place
 - Distractions during the assessment

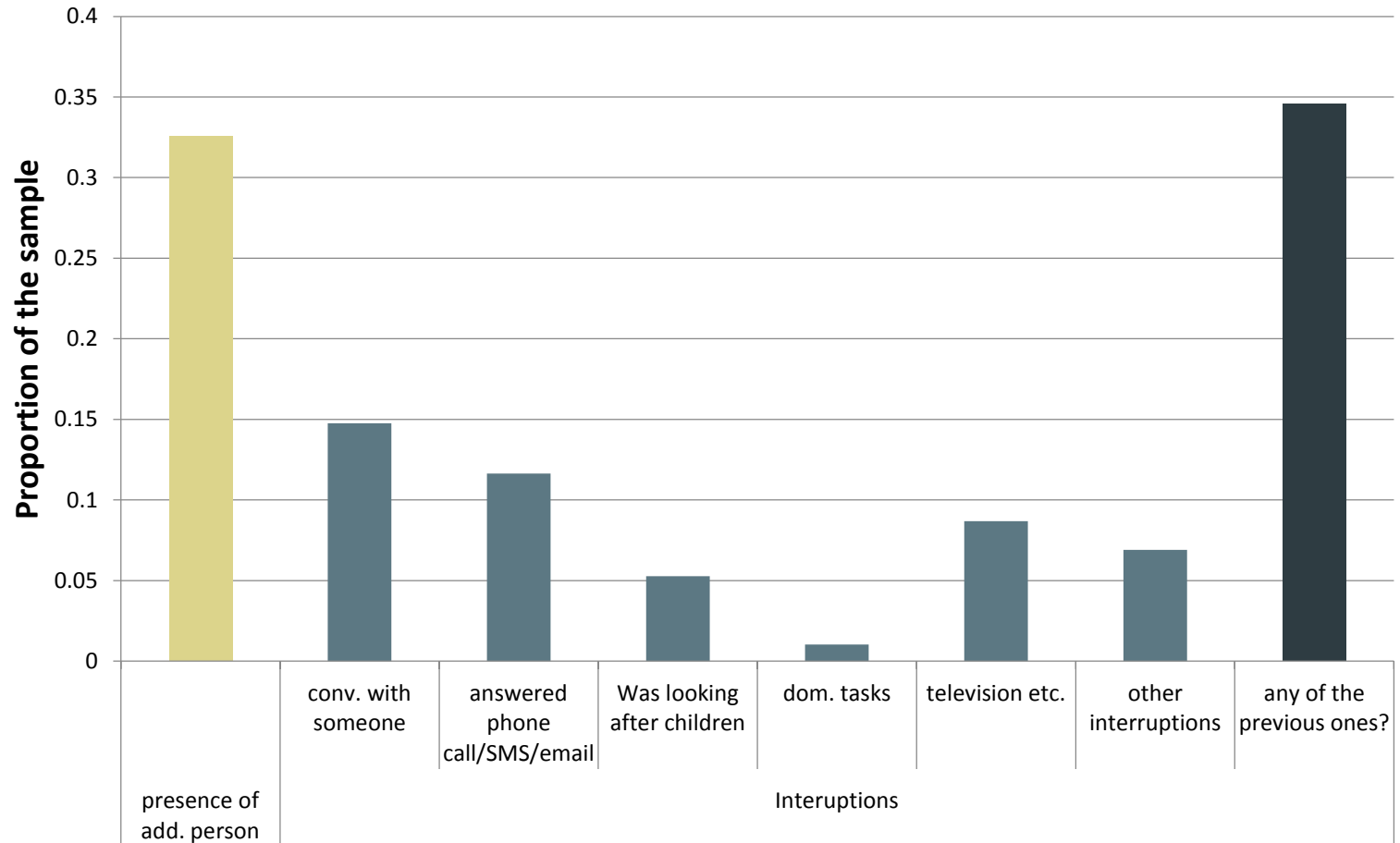


Some analysis of interviewer effects



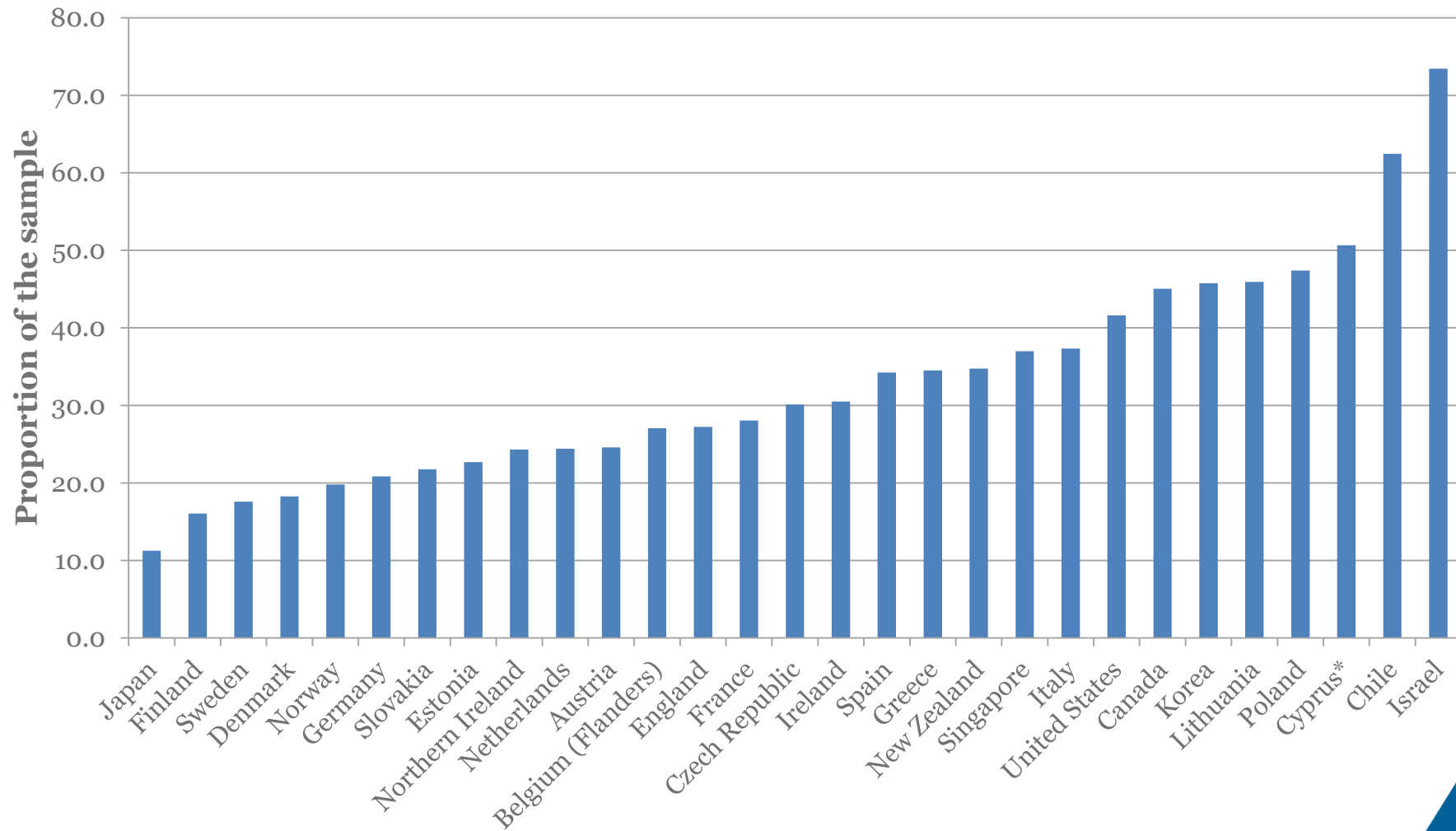


Interruptions: country averages



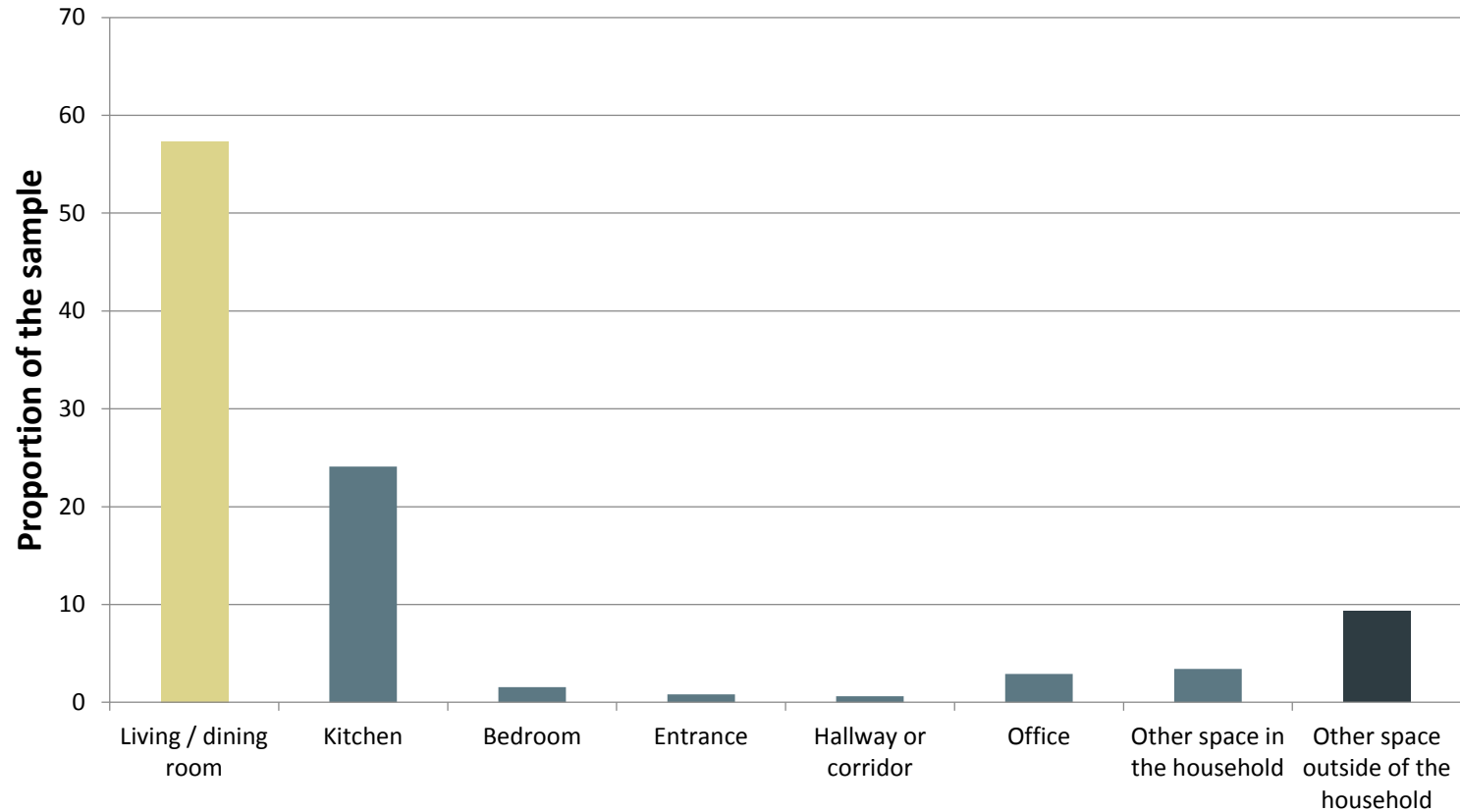


Other person present: variation across countries



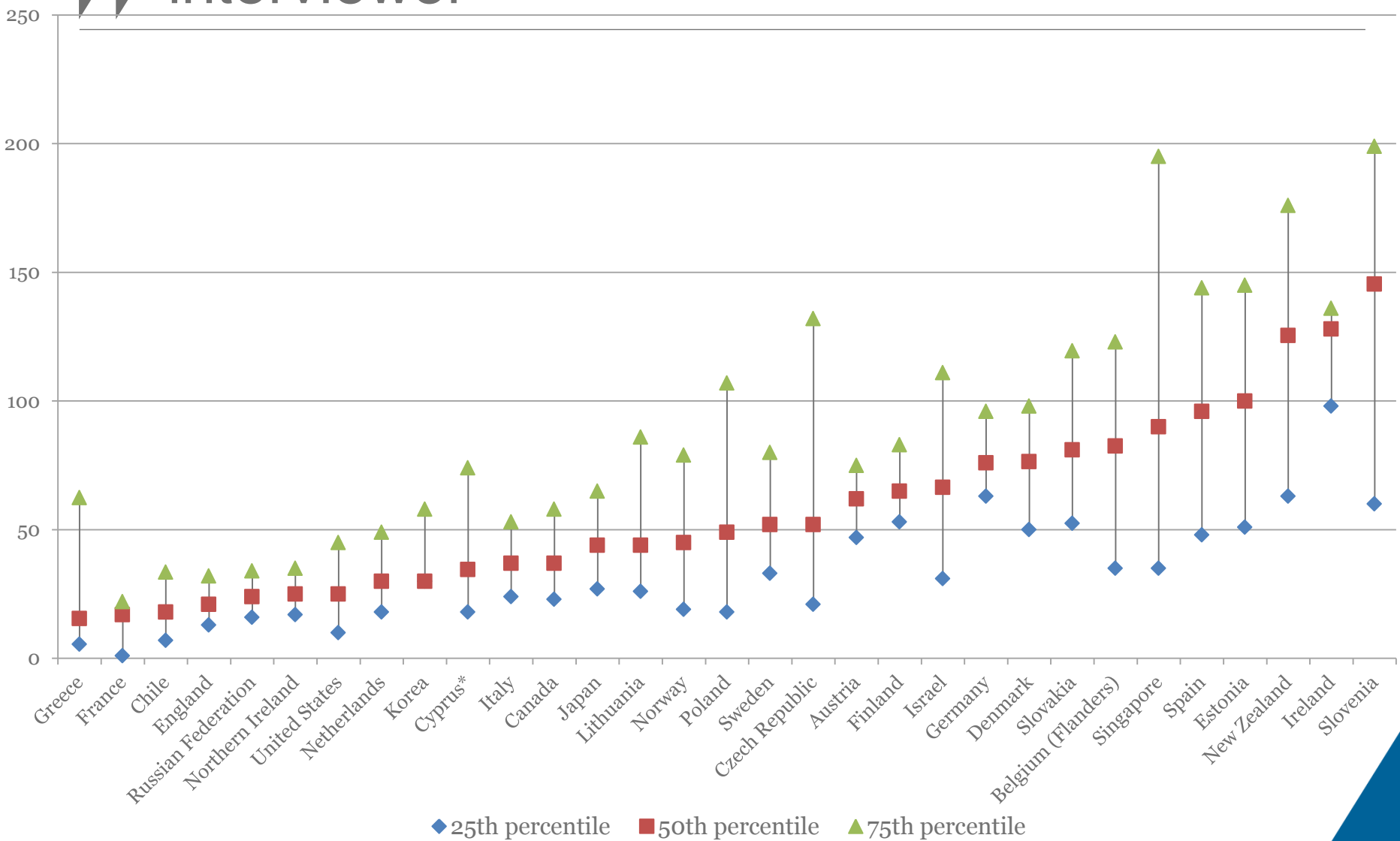


Room of assessment: country averages



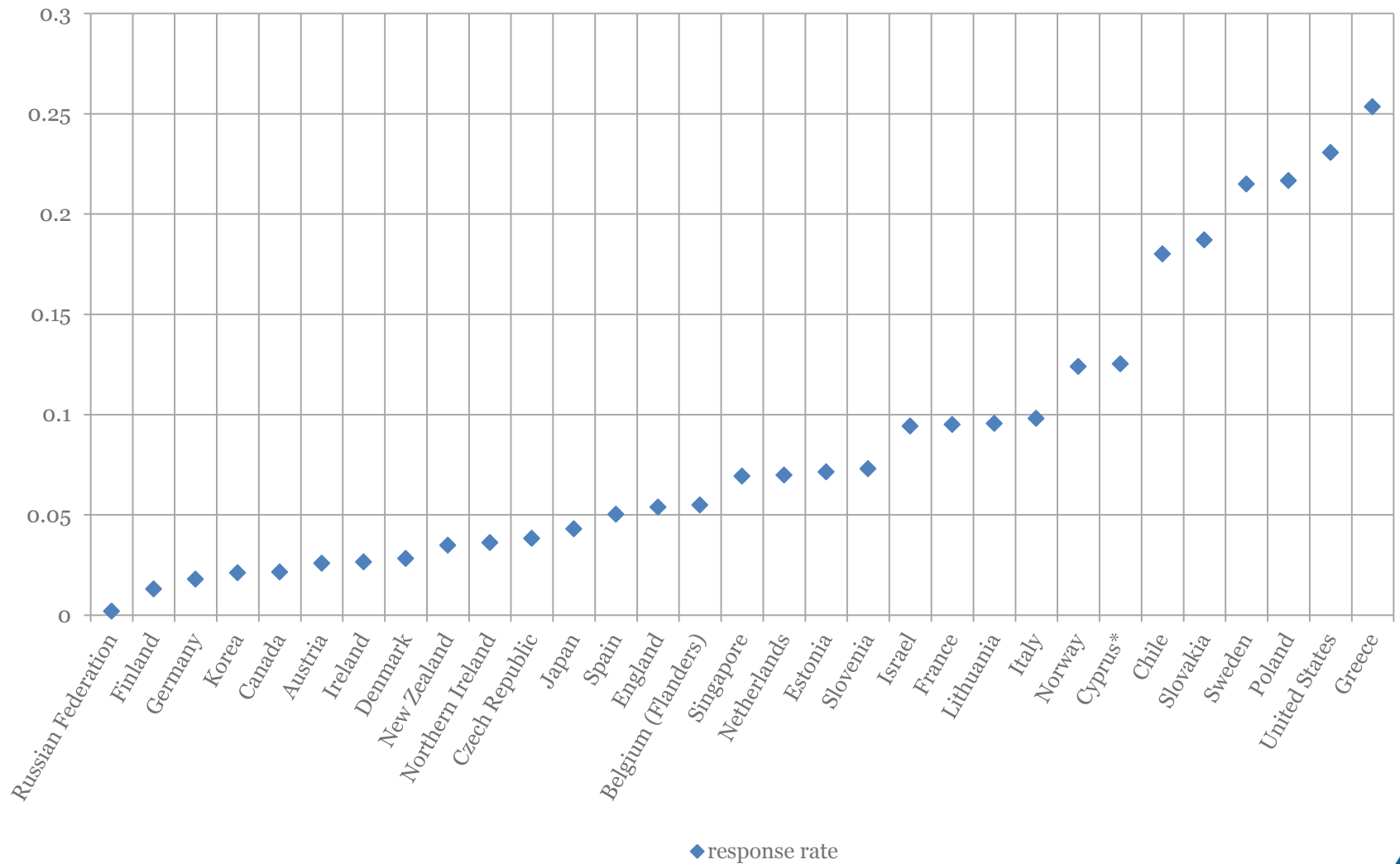


Distribution of number of assignments by interviewer



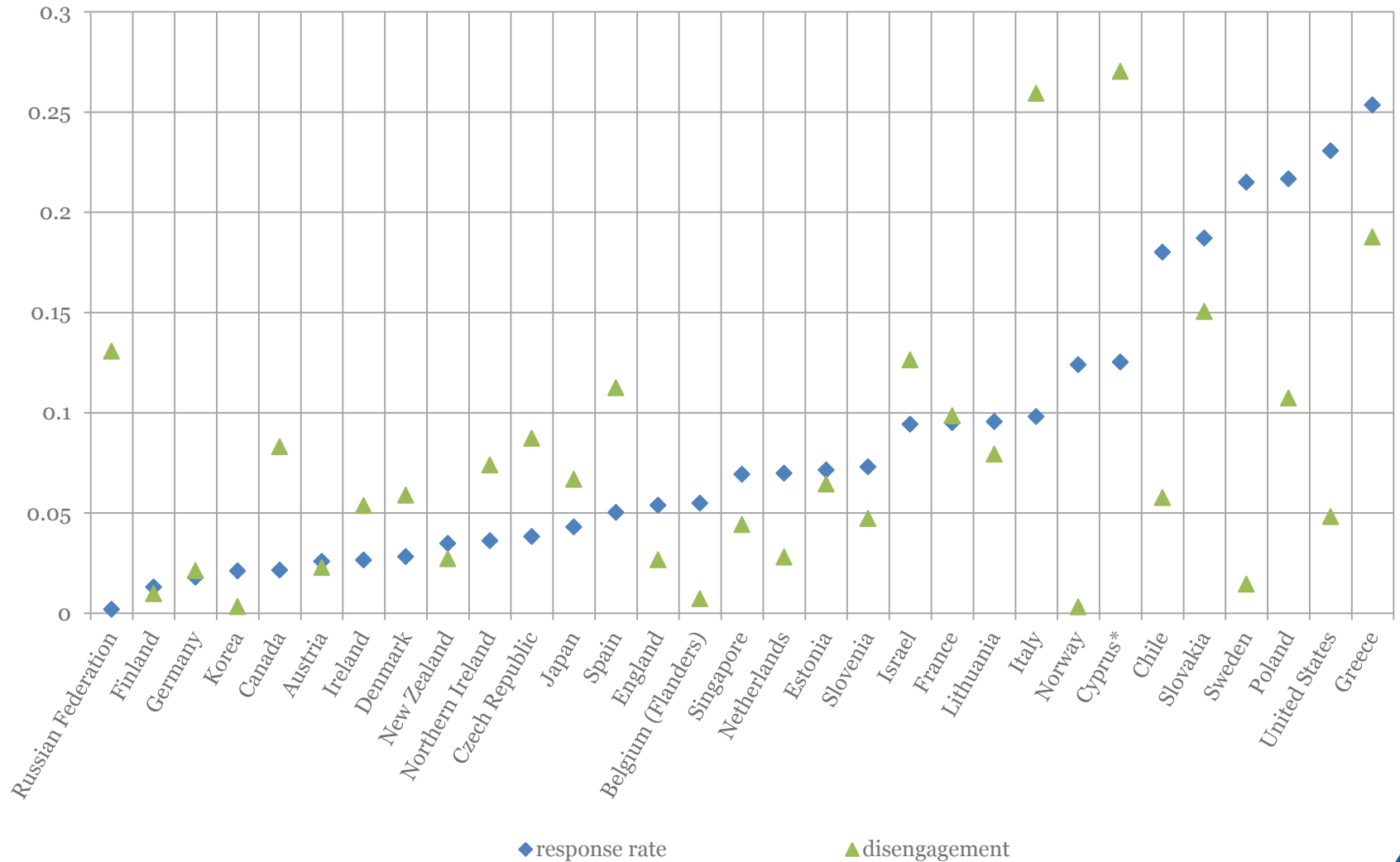


Interviewer intra-correlation in PIAAC





Interviewer intra-correlation in PIAAC





Understanding the Interview Process

Roger Tourangeau:

- Cognitive processes needed to give a correct answer
- CASM – comprehension, retrieval, judgement, reporting
- Alternatives (briefly touched upon)
 - Satisficing
 - Conversational analysis

Bryan Maddox:

- Interviewer/respondent/computer interaction
- Home is a complex testing situation
- Importance of interviewer in ensuring respondent displays competence



Data Fabrication

Kentaro Yamamoto:

- Examples of data fabrication in PIAAC and PISA
- PIAAC
 - 1st example: duplicate cases, highly prolific interviewers with unusually high proficiency and low variation among respondents.
 - 2nd example: prolific interviewers with fast response times and high proportions of omitted cognitive items for respondents
- PISA – (inflation of scores by coders)
- Detection and remediation of falsification
- Recommendation: systematic approach to the identification of fabrication and that this is undertaken closer to the timing of data collection

Jörg Blasius:

- Data fabrication is a form of task simplification
- Analysis of PISA and PIAAC BQs
- Principal Components and multiple correspondence analysis used to identify UDPs and IRPs
- The methods should be applied during the field period where suspicious interviews can be detected (and checked).



Dealing with Survey Error

Geert Loosveldt:

- Evaluation of interviewer effects is part of survey data quality assessment
- Two approaches
 - Interviewer behavior and related measurement errors
 - Interviewer variance analysis
- ICCs (intra-interviewer correlations):
- 2 examples (behavioural and attitudinal variables relating to alcohol use)
- Interviewer effects have impact on substantive conclusions (parameters St errors and correlations)

Matthias von Davier:

- Overview of statistical methods that can be used to detect sources of error in assessments and questionnaires
- Assumptions: Item homogeneity and person homogeneity
- Differential item fit across populations
- Response styles
- BQ and response styles
- Warning: better to regroup and look at what we have in our assessment that help to understand response behaviour before inventing and collecting data on new skills



Detection of Errors during the Interview

Mohadjer & Edwards:

- Describe the use of survey dashboards
- Evolution of dashboards in use at Westat
- Example of PIAAC
- Conclusions : real time access to key indicators can improve data quality
 - PIAAC should invest in CMS, CARI and timing data

Hibben, Pennell, Scott:

- Interviewer error has big effects in 3MC studies
- Examples of QC/QA approaches to minimise interviewer error
 - Validation in real time
 - GPS
- Recommendations for PIAAC
 - Contracting
 - Interviewer recruitment and training
 - Data collection



The Future of Data Collection

Michael Schober:

- Examines the future of FTF interviewing
- Trends that point to lesser reliance on such techniques
 - New modes of communication
 - Falling response rates
 - high costs
- More evidence on advantages and disadvantages of different modes. No longer obvious the FTF is gold standard
- FTF will continue but pressures to complement it with other contact modes.

Michael Link:

- Reviews new data collection strategies to overcome challenges to traditional modes
- Non-probability sampling
 - Interest growing, range of techniques, but probability sample still uses where reliable and valid population estimates are required. Situation may change
- Mobile data capture (use of mobile devices tablets, smartphones, etc.)
 - Lots of experimentation, but not necessarily the case that public is ready for these as reliable data collection tools
- Big Data
 - Promises much but faces problems similar to non-probability sampling
- Each area shows promise but often severe shortcomings



Some themes/reflections

1. Has been interesting to get psychometricians and survey methodologists together
2. What occurs in the interview situation
 - Interviewer as ally or source of error
3. ICT
 - Opportunities and challenges (more data, more possibilities for collection, many questions)
 - Methodological revolution?
 - How to harness possibilities without sacrificing representation and rigour?



Implications (1): for PIAAC

- 2nd cycle of PIAAC
 - 2018-2023
 - Opportunity to think more deeply about monitoring of data collection, identification of problems, possible treatments than was possible in the 1st Cycle.



Implications

- Development of monitoring tools for the 2nd cycle
 - More complete and timely information may have helped avoid some of the problems encountered in the 1st cycle
 - Expectation that the contractor will develop tools and indicators that exploit available paradata
 - Dashboard approach presented by Leyla Mohadjer is obviously of great interest as are the examples offered by Beth-Ellen Pennell



Implications

- Technical Standards and Guidelines
 - Review in the light of best practice regarding control monitoring of field operations
 - Draw from other studies
 - Suggestions from Beth-Ellen Pennell useful
- Will need to explore possibilities of multi-modal approach to data collection



Implications

- Be more systematic in collecting information regarding interviewers and make it available for analysis
 - Include interviewer data in PUFs
 - Privacy and data protection legislation sets some limits
- Systematically review data to identify fabrication, falsification
 - Many techniques available
 - Should be undertaken close to data collection



Further research

- The interviewer/respondent relationship
- Testing conditions and performance
- Response styles
- Motivation of test-takers
- Alternatives data collection methods relevant to large scale testing

Find Out More About PIAAC at:



www.oecd.org/skills/piaac

All national and international
publications

The complete micro-level database



Email

william.thorn@OECD.org

Thank you