

# Addressing Outcome-Related Nonresponse Through a Doorstep Interview

Wendy Van de Kerckhove<sup>1</sup>, Tom Krenzke<sup>1</sup>, Leyla Mohadjer<sup>1</sup>

<sup>1</sup>Westat, 1600 Research Blvd, Rockville, MD 20850

## Abstract

Not missing-at-random (NMAR) nonresponse occurs when the reason for not responding to a survey directly relates to the outcome of interest. This source of bias cannot be addressed through standard weighting adjustments that rely on auxiliary variables. The Programme for the International Assessment of Adult Competencies (PIAAC) administers adult skills assessments (e.g., literacy, numeracy) in a country's main language(s). It is subject to NMAR nonresponse because of persons that are unable to participate due to a language barrier, learning/mental disability, or reading/writing barrier. This paper provides background on the handling of such cases in data collection, weighting, and estimation for the first cycle of PIAAC. It then describes plans to further reduce the extent of potential bias due to NMAR nonresponse in the second cycle by introducing a brief multi-lingual questionnaire called the Doorstep Interview.

**Key Words:** non-ignorable, not missing-at-random (NMAR), informative, bias, PIAAC

## 1. Introduction

Surveys are subject to not missing-at-random (NMAR) nonresponse when the reason for not responding to the survey is related to the survey outcome, even after accounting for auxiliary characteristics. This type of nonresponse is also referred to as non-ignorable or informative. It cannot be addressed through standard weighting adjustments that rely on a missing-at-random assumption, and so it can cause biased survey estimates.

The Programme for the International Assessment of Adult Competencies (PIAAC) is an international survey of adults ages 16 to 65, sponsored by the Organisation for Cooperation and Development (OECD), which assesses proficiency in the language(s) of the country through an in-person household survey. It is subject to NMAR nonresponse because of adults who are unable to complete the survey due to a language barrier, reading/writing difficulty, or learning/mental disability. The PIAAC Consortium has implemented various procedures to address this source of nonresponse and limit the associated nonresponse bias.

### 1.1 NMAR Nonresponse

Standard procedures for reducing nonresponse bias through weighting or imputation, such as those described in Kalton and Flores Cervantes (2003) or Kalton and Kasprzyk (1986), are based on the assumption that the missing data is missing-at-random. In other words, the probability of nonresponse is independent of the survey outcome after controlling for auxiliary characteristics. Given the presence of NMAR nonresponse in PIAAC, the standard weighting and imputation procedures are not appropriate.

Several papers evaluate alternative methods for addressing nonresponse when the missing-at-random assumption does not hold. For example, Kott and Chang (2010) describe the use of calibration weighting, Pfeffermann and Feder (2013) take an empirical likelihood approach, and Siddique and Belin (2008) address NMAR in the multiple imputation framework with approximate Bayesian bootstrap. Another example is Laaksonen and Chambers (2006), who evaluated using a nonresponse follow-up survey to collect some key outcomes and utilizing that data in the weighting or prediction models. In PIAAC, we plan to implement an additional data collection effort for the NMAR nonrespondents. The data will be collected through a Doorstep Interview, and we will use the association between the reason for nonresponse and the outcome, along with the Doorstep Interview data, in the weighting and estimation process to reduce the bias associated with NMAR nonresponse.

## 1.2 PIAAC

The first cycle of PIAAC occurred over three rounds of data collection, in 2011-2012, 2014-2015, and 2017-2018, with 39 countries participating. Thirty-three countries are planning to take part in the second cycle, with the main data collection scheduled to occur in 2022 and a field trial in 2021. The majority of countries in the second cycle also participated in the first cycle.

In PIAAC, a background questionnaire (BQ) and an assessment are administered. The BQ collects demographics and other background information. The assessment evaluates literacy, numeracy, and problem-solving skills. Countries that do not have a population registry also administer a screener to identify and sample eligible persons within a household prior to the BQ.

The key outcomes for PIAAC consist of proficiency estimates for literacy, numeracy, and problem solving. Because each respondent does not receive the same set of assessment items, a proficiency score for each respondent cannot be produced by simply calculating the number of correct responses. Instead, item response theory and latent regression modeling are applied using the assessment responses and the BQ data. Multiply imputed proficiency values (plausible values or PVs) are generated to account for the error at the individual-level. A more detailed description of the process can be found in chapter 17 of the PIAAC Technical Report (Yamamoto et al., 2019).

The responsibility for survey administration is divided between the countries and the Consortium. Countries are responsible for sampling and data collection, with guidance from the Consortium. The Consortium (Westat) is responsible for weighting, although countries can choose to perform this task, especially if there are confidentiality concerns. The Consortium (Educational Testing Service [ETS]) generates the PV's for the proficiency estimates. In addition, the Consortium produces Technical Standards and Guidelines that countries and the Consortium must adhere to throughout the survey process, along with a series of quality control checks. The goal is to ensure high-quality data that is comparable across countries.

## 2. Motivation for the Doorstep Interview

Adults that cannot complete the BQ for a literacy-related reason are part of the PIAAC target population but do not have sufficient data to generate or impute scores. In addition, they cannot be represented by respondents in weighting adjustments because they are known to have lower proficiency even after accounting for auxiliary information. This

source of NMAR nonresponse needs to be addressed or it could result in biased proficiency estimates.

### **2.1 NMAR Approach in Cycle 1**

In PIAAC Cycle 1, the Consortium developed strategies to address NMAR nonresponse throughout the survey process. During data collection, countries were encouraged to translate the screener (if applicable) and BQ into multiple languages, and to use bilingual interviewers or interpreters in administering the screener and BQ when needed. The Consortium developed specific disposition codes to identify cases that could not complete the survey for literacy-related reasons. In addition, the weighting standards required that the NMAR cases were separated out in nonresponse adjustments, and both BQ respondents and BQ NMAR non-respondents received a final weight. Van de Kerckhove, Mohadjer, and Krenzke (2013) provide further details on the treatment of NMAR cases in the weighting process.

For the purpose of estimation, the Consortium (ETS) imputed scores for sampled persons that completed the BQ but not the assessment. The process is described in Yamamoto et al. (2019). To address NMAR nonresponse at the assessment stage, the Consortium imputed incorrect responses to assessment items for those that could not complete the assessment for a literacy-related reason (language barrier, reading/writing barrier, or learning/mental disability). Scores were generated for such cases using the incorrect responses and the BQ data. For other BQ respondents without assessment data, assessment items were treated as missing and scores imputed using BQ data. NMAR nonrespondents to the BQ could not be treated in the same manner as those at the assessment stage because they lacked sufficient background data to generate scores. Therefore, the BQ NMAR cases have a final weight but no proficiency scores.

Figures 1 and 2 provide examples of the reporting of proficiency estimates for PIAAC. Figure 1 shows mean literacy proficiency scores for all countries. Given that the BQ NMAR cases lack scores, they are excluded from means and are not considered in the country rankings. The exclusion is noted in a footnote to the figure. Figure 2 shows the percentage of the population in each proficiency level for literacy, where a description of the levels can be found in OECD (2019). In this figure, the BQ NMAR population is given as a separate “missing” category, with an explanation provided in the footnote.

Mean literacy proficiency scores of 16-65 year-olds

Mean	Comparison country	Countries whose mean score is NOT significantly different from the comparison country
296	Japan	
288	Finland	
284	Netherlands	
281	New Zealand	Australia, Sweden, Russian Federation <sup>1</sup>
280	Australia	New Zealand, Norway, Sweden, Russian Federation <sup>1</sup>
279	Sweden	Australia, New Zealand, Norway, Russian Federation <sup>1</sup>
278	Norway	Australia, Sweden, Russian Federation <sup>1</sup>
276	Estonia	Czech Republic, Flanders (Belgium), Russian Federation <sup>1</sup>
275	Flanders (Belgium)	Czech Republic, Estonia, Slovak Republic, Russian Federation <sup>1</sup>
275	Russian Federation <sup>1</sup>	Australia, Canada, Czech Republic, Denmark, England (UK), Estonia, Flanders (Belgium), Germany, Korea, New Zealand, Northern Ireland (UK), Norway, Slovak Republic, Sweden, United States 2012/2014, United States 2017
274	Czech Republic	Canada, England (UK), Estonia, Flanders (Belgium), Korea, Slovak Republic, United States 2012/2014, United States 2017, Russian Federation <sup>1</sup>
274	Slovak Republic	Canada, Czech Republic, England (UK), Flanders (Belgium), Korea, United States 2012/2014, Russian Federation <sup>1</sup>
273	Canada	Czech Republic, England (UK), Korea, Slovak Republic, United States 2012/2014, United States 2017, Russian Federation <sup>1</sup>
273	England (UK)	Canada, Czech Republic, Denmark, Korea, Northern Ireland (UK), Slovak Republic, United States 2012/2014, United States 2017, Russian Federation <sup>1</sup>
273	Korea	Canada, Czech Republic, England (UK), Northern Ireland (UK), Slovak Republic, United States 2012/2014, United States 2017, Russian Federation <sup>1</sup>
272	United States 2012/2014	Austria, Canada, Czech Republic, Denmark, England (UK), Germany, Korea, Northern Ireland (UK), Slovak Republic, United States 2017, Russian Federation <sup>1</sup>
271	United States 2017	Austria, Canada, Czech Republic, Denmark, England (UK), Germany, Korea, Northern Ireland (UK), United States 2012/2014, Cyprus <sup>2</sup> , Russian Federation <sup>1</sup>
271	Denmark	Austria, England (UK), Germany, Northern Ireland (UK), United States 2012/2014, United States 2017, Russian Federation <sup>1</sup>
270	Germany	Austria, Denmark, Northern Ireland (UK), United States 2012/2014, United States 2017, Cyprus <sup>2</sup> , Russian Federation <sup>1</sup>
269	Austria	Denmark, Germany, Northern Ireland (UK), United States 2012/2014, United States 2017, Cyprus <sup>2</sup>
269	Cyprus <sup>2</sup>	Austria, Germany, Ireland, Lithuania, Northern Ireland (UK), United States 2017
269	Northern Ireland (UK)	Austria, Denmark, England (UK), Germany, Ireland, Korea, Lithuania, Poland, United States 2012/2014, United States 2017, Cyprus <sup>2</sup> , Russian Federation <sup>1</sup>
267	Poland	Ireland, Lithuania, Northern Ireland (UK)
267	Lithuania	Ireland, Northern Ireland (UK), Poland, Cyprus <sup>2</sup>
267	Ireland	Lithuania, Northern Ireland (UK), Poland, Cyprus <sup>2</sup>
266	OECD average	Ireland, Lithuania, Northern Ireland (UK), Poland
264	Hungary	France
262	France	Hungary
258	Singapore	Slovenia
256	Slovenia	Greece, Israel, Singapore
255	Israel	Greece, Slovenia
254	Greece	Israel, Slovenia, Spain
252	Spain	Greece, Italy
250	Italy	Spain, Kazakhstan
249	Kazakhstan	Italy
227	Turkey	
222	Mexico	Chile
220	Chile	Mexico
196	Ecuador	Peru
196	Peru	Ecuador

Note: Statistical significance is at the 5% level. Literacy-related non-response (missing) is excluded from the calculation of mean scores.

1. See note at the end of this chapter.

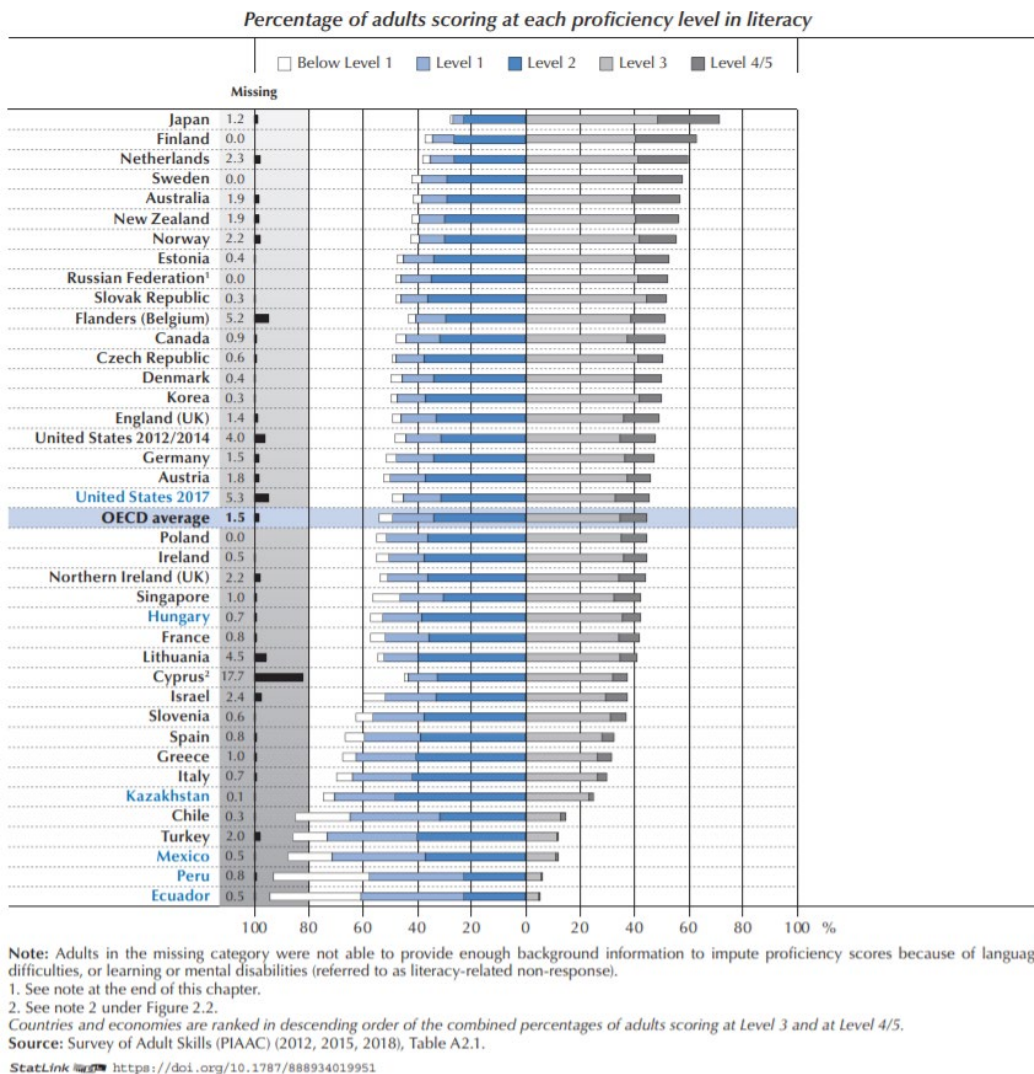
2. See note 2 under Figure 2.2.

Countries and economies are ranked in descending order of the mean score.

Source: Survey of Adult Skills (PIAAC) (2012, 2015, 2018), Table A2.2.

StatLink  <https://doi.org/10.1787/888934019970>

Figure 1: Reporting of PIAAC mean literacy scores, Figure 2.7 from OECD (2019)



**Figure 2:** Reporting of PIAAC literacy levels, Figure 2.6 from OECD (2019)

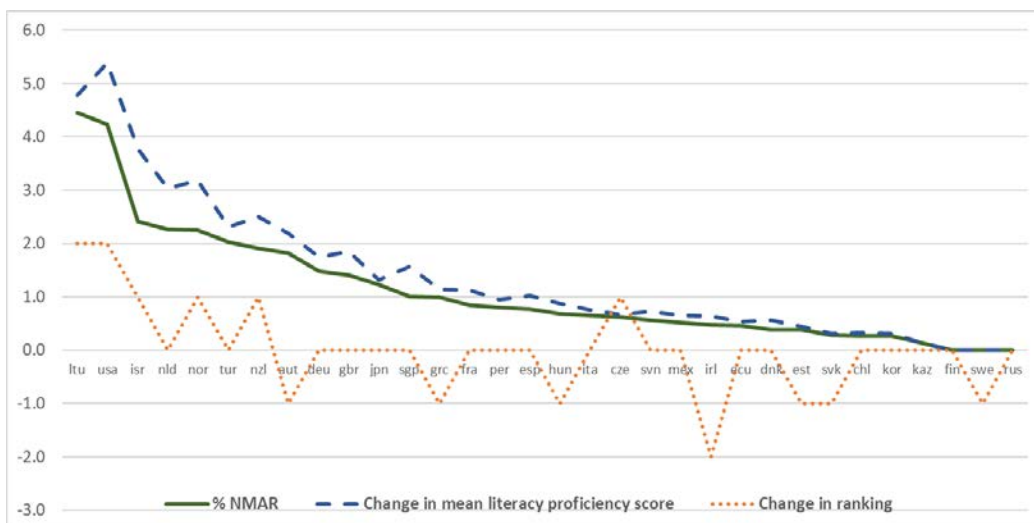
### 2.2 Potential Effect on Bias

Including BQ NMAR cases on data files with final weights allows estimation of the proportion of the population meeting this criterion. However, the missing scores mean that this group cannot be included in mean proficiency estimates and will likely be ignored by analysts. Ignoring the BQ NMAR cases can lead to biased estimates.

Using the PIAAC public use data for 32 countries<sup>1</sup>, we evaluated the potential bias from excluding the BQ NMAR cases from mean literacy proficiency estimates. For this analysis, we assumed that the BQ NMAR cases would have low literacy scores – at the first

<sup>1</sup> Public use data is available for 35 countries. Three countries - Belgium (Flanders), Canada, and Poland - were excluded from our analysis. Unlike other countries, Belgium’s cases designated as language barriers are believed to be veiled refusals that do not necessarily have low proficiency in the assessment language. Canada’s BQ NMAR cases are not provided in the public use file for disclosure reasons. Poland’s BQ NMAR cases were treated as MAR nonrespondents and were not assigned final weights, deviating from the weighting standards.

percentile<sup>2</sup> of the scores for respondents. We then compared the mean scores with and without the BQ NMAR cases. The results based on the first percentile assumption are provided in Figure 3. The bias in the mean literacy proficiency score is under 6 points for all countries, and a country’s ranking would change by at most two. For example, for Lithuania (LTU), NMAR cases are 4.5 percent of the target population. Based on the previously stated assumptions, excluding the NMAR cases results in an overestimate of the mean literacy proficiency score by about 5 points, and Lithuania’s ranking is two places higher than if the NMAR were included. In all countries in the analysis, under 5.0 percent of the population was unable to complete the BQ for a literacy-related reason (NMAR), and so excluding this population is associated with only a slight nonresponse bias under the current assumptions.



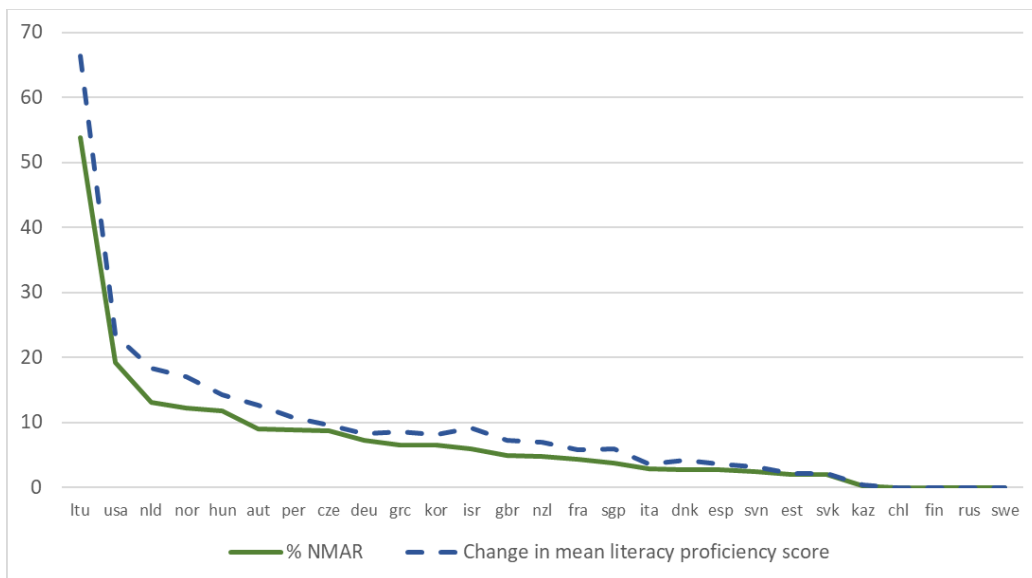
**Figure 3:** Effect of excluding non-ignorable (NMAR) nonrespondents from PIAAC literacy score averages, assuming NMAR nonrespondents would score at the first percentile of scores for respondents

While the potential bias in national estimates is small given the low proportions of NMAR, the effect could be larger for certain subgroups. Figure 4 shows results from repeating the above evaluation for the foreign-born population in 27 countries.<sup>3</sup> For BQ respondents, foreign-born status is based on the variable J\_Q04A from the BQ. For BQ NMAR, we assumed that all those with language barriers are foreign born, and those with a reading/writing barrier or learning/mental disability are native born. Under the assumption that the NMAR cases would score at the first percentile of scores for respondents, the bias could be substantial – the mean would be overestimated by 10 or more points for eight countries, with a 23-point difference for the United States and 66-point difference for Lithuania.

<sup>2</sup> The analysis was also performed assuming NMAR cases would score at the fifth percentile of respondents’ scores, but the results were similar and are not shown here.

<sup>3</sup> Five additional countries are excluded from this sub-group analysis. Ireland’s data file did not include the specific reason for NMAR nonresponse. Ecuador, Japan, Mexico, and Turkey had less than 30 foreign-born cases.





**Figure 4:** Effect of excluding non-ignorable (NMAR) nonrespondents from PIAAC literacy score averages for the foreign-born population, assuming persons with language barriers are foreign-born and would score at the first percentile of scores for respondents

### 3. Introduction of the Doorstep Interview

To minimize the bias associated with NMAR nonresponse, the PIAAC Consortium introduced a new component to the survey process for Cycle 2. The new component is a shortened, multi-lingual version of the BQ called the Doorstep Interview. The Doorstep Interview will be administered when the sampled person is unable to complete the full BQ because of a language barrier. It contains key background items on gender, age, education, employment status, and native/non-native status. These items, along with the reason for nonresponse, will provide sufficient data to impute scores.

Through combined efforts of the participating countries and the Consortium, the Doorstep Interview has been translated into over 40 languages, and each country can determine which languages to offer. The process for the Doorstep Interview involves three main steps:

- First, if the interviewer and respondent do not have a shared language, then the interviewer will show the respondent the **Language Identification card**. An illustration of the card is provided in Figure 5.
- After the respondent indicates their preferred language, the interviewer will show them the **Doorstep Introduction card** in the corresponding language. This card introduces the study and asks for the sampled person.
- Finally, if the sampled person does not speak the BQ language(s) and there is no interpreter available, then the sampled person will complete the **Doorstep Interview module**. The module will be self-administered on the tablet in the preferred language. As the name implies, it can literally be completed at the doorstep.

01	Point to the language you best understand
02	Señale el idioma que entienda mejor
03	選擇您最能夠掌握的語言
04	Thể Giới thiệu Ngôn ngữ

**Figure 5:** Illustration of a PIAAC Language Identification card

For quality control, the Consortium will closely monitor the number of Doorstep Interviews completed in each country. Prior to data collection, each country has provided the Consortium with an estimate of the percentage of cases expected to be routed toward the Doorstep Interview based on experience from PIAAC Cycle 1 and/or national population statistics on languages spoken in the country. During data collection, countries will provide monthly reports to the Consortium showing the actual percentage of Doorstep Interview cases. If this substantially exceeds expectations, then countries will need to investigate the reason and ensure that it is not a result of interviewers favoring this lower level-of-effort option.

With the completion of the Doorstep Interview, the BQ NMAR cases can be treated in a similar manner as the assessment NMAR cases in estimation and reporting. The Consortium can impute incorrect responses to the assessment items and generate scores using the incorrect responses and Doorstep Interview data. This will allow BQ NMAR cases to be included in final proficiency estimates.

#### 4. Discussion

The introduction of the Doorstep Interview in PIAAC Cycle 2 is intended to reduce bias in proficiency estimates. The brief, multi-lingual questionnaire will allow key background data to be collected for persons with language barriers that are unable to complete the full BQ. As a result, the Consortium will be able to impute scores for these NMAR nonrespondents by using the data from the Doorstep Interview and the knowledge that such respondents would have scored low on the assessment because of the language barrier. This is an improvement over Cycle 1, in which analysts were able to estimate the proportion of the population that could not complete the BQ for a literacy-related reason but were not able to include such cases in the estimates of mean proficiency scores.

Potential challenges with the Doorstep Interview relate to cost, scope, and trends. First, including this additional component comes with costs for instrument development, translation, and training. For countries that have a very low prevalence of NMAR cases, the Doorstep Interview will have minimal effect on their national estimates, and so they could feel that the cost outweighs the benefit. However, one goal of PIAAC is to produce comparable estimates across countries. While the Doorstep Interview might not have a large effect on a particular country's estimates, it could have a larger effect on the countries with which it is being compared. It is important to implement the survey in a consistent



manner across all countries. In addition, the proportion of NMAR cases could be larger than expected because of recent immigration changes.

The second potential challenge is that the Doorstep Interview is limited in scope because it applies only to persons that cannot complete the BQ because of a language barrier. It does not allow the collection of data for persons with reading/writing barriers or learning/mental disabilities. In addition, the sampled person must be identified before the Doorstep Interview can be administered. Therefore, countries that require a screener for the purpose of sample selection will first need to complete the screener with a household respondent. However, for most countries in Cycle 1, the majority of NMAR nonresponse occurred at the BQ stage and was usually attributable to language barriers. In addition, the Consortium plans to address the limited scope of the Doorstep Interview by allowing the assessment NMAR cases and the Doorstep Interview respondents to represent other NMAR cases in the nonresponse weighting adjustments. This is based on the assumption that the NMAR cases that complete the Doorstep Interview or assessment would have similar proficiency estimates as other NMAR cases. In Cycle 1, most countries did not have a sufficient number of NMAR cases at the assessment stage to represent those at earlier stages, but we are anticipating that sample sizes will be sufficient with the addition of Doorstep Interview respondents.

Finally, with the inclusion of BQ NMAR cases, the Cycle 2 proficiency estimates will not be directly comparable with the Cycle 1 estimates. However, it is still possible to evaluate trends. For analysis of proficiency levels, the BQ NMAR cases can be assumed to be at or below Level 1. For analysis of means, one option is to perform a domain analysis, comparing the non-NMAR cases from Cycle 1 to those in Cycle 2. For the countries that had a sufficient number of NMAR cases at the assessment stage, another option is to impute scores for the BQ NMAR cases using the assessment NMAR cases. Other countries could consider a model-assisted approach that builds strength from the association between Cycle 1 and Cycle 2 observed values.

The Doorstep Interview will be tested in the PIAAC Cycle 2 field trial, which is currently delayed until the spring of 2021.

### References

- Kalton, G., and I. Flores-Cervantes. 2003. Weighting Methods. *Journal of Official Statistics*. 19(2), 81-97.
- Kalton, G., and D. Kasprczyk. 1986. The Treatment of Missing Survey Data. *Survey Methodology*. 12, 1-16.
- Kott, P.S., and T. Chang. 2010. Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association*. 105(491), 1265-1275.
- Laaksonen, S., and R. L. Chambers. 2006. Survey estimation under informative nonresponse with follow-up. *Journal of Official Statistics: an international quarterly*. 22(1), 81-95.
- OECD. 2019. *Skills matter: Additional results from the Survey of Adult Skills, OECD Skills Studies*. OECD Publishing. Paris. Retrieved July 2, 2020, from [https://www.oecd.org/skills/piaac/publications/Skills\\_Matter\\_Additional\\_Results\\_from\\_the\\_Survey\\_of\\_Adult\\_Skills\\_ENG.pdf](https://www.oecd.org/skills/piaac/publications/Skills_Matter_Additional_Results_from_the_Survey_of_Adult_Skills_ENG.pdf)
- Pfeffermann, D., and M. Feder. 2013. Fitting models to complex survey data accounting for nonignorable sampling and nonresponse. *Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS087)*. 894-899.

- Siddique, J., and T.R. Belin. 2008. Using an approximate Bayesian bootstrap to multiply impute nonignorable missing data. *Computational Statistics & Data Analysis*. 53(2), 405-415.
- Van de Kerckhove, W., L. Mohadjer, and T. Krenzke. 2013. Treatment of Outcome-Related Nonresponse in an International Literacy Survey. *JSM Proceedings, Survey Research Methods Section*. Alexandria, VA: American Statistical Association. 1328-1336.
- Yamamoto, K., L. Khorramdel, M. von Davier, U.S. Ali, and F. Robin. 2019. Scaling PIAAC Cognitive Data. *Technical Report of the Survey of Adult Skills (PIAAC) (Third Edition)*, section 5. OECD. Retrieved July 2, 2020, from [https://www.oecd.org/skills/piaac/publications/PIAAC\\_Technical\\_Report\\_3rd\\_Edition\\_2019\\_Section5\\_Chapters17-23.pdf](https://www.oecd.org/skills/piaac/publications/PIAAC_Technical_Report_3rd_Edition_2019_Section5_Chapters17-23.pdf)