An examination of the identification problem arising from unit nonresponse in ILSA studies

Report

Diego Cortes, Jeff Dominitz, Maximiliano Romero, Sabine Meinck

November 2023

1. Introduction

The purpose of this report is to inform and provide recommendations to the IEA and its technical expert group on technical standards and reporting, based on results obtained in the project *Studying unit nonresponse in ILSA studies using the partial identification framework*. This report uses results derived in its accompanying manuscript and therefore should be read together with it.

The technical standards for IEA surveys recognize that, when estimating population parameters with survey data, participation rates provide "quantitative information (…) to indicate the potential for non-sampling error" (Martin, 1999, p.71). Accordingly, in the technical report of each IEA study, there is a section in which participation rates are made available to the reader. For example, in the International Computer and Information Literacy Study (ICILS), weighted and unweighted participation rates are reported to 'facilitate the evaluation of data quality and reduce the risk of potential bias due to non-response' (Tieck, 2020, p. 84). Put together, we understand that the IEA regards participation rates as a measure that quantifies the risk of bias in estimation introduced by nonresponse. The lower the participation rate is in a study, the higher is the risk of potential bias, and vice versa.

IEA's strategy to reflect on this risk of bias in the inference is to categorize study populations according to their achieved participation rate.[1] Overall, there are three categories: satisfactory (category 1), satisfactory only after the inclusion of replacement schools (category 2), and unsatisfactory (category 3). The rules for categorizing participation rates are reported in the Technical Report of each study. Results from study populations reaching participation rates of *category 1* are reported without any annotation; results from study populations reaching participation rates of *category 2* are reported with a cautionary annotation; while results from study populations reaching participation rates of *category 3* are reported in separate tables. Overall, the motivation of this strategy is to signal the reader that the data collected in a study population falling in *category 1* leads to credible inferences, while the credibility of the inference in study populations reaching only *category 3* is, to a large extent, questionable.

The argument presented above focuses attention solely on participation rates and makes no explicit statement about distributional assumptions of the outcome of interest among nonparticipating units. In the accompanying manuscript, a thorough examination of this issue is made using the partial identification framework (Manski 1995; Manski, 2003). This framework makes intensive use of the Law of Total Probability to provide clarity about what can be learned from the probability

[1] To clarify terminology, in this report we use the term study population to refer to each entity that participates and is reported separately in an IEA study.

distribution of interest $P(y)$ with the data collected, provided that some part of the population is not participating in the survey. That is,

$$P(y) = P(y|z = 1)P(z = 1) + P(y|z = 0)P(z = 0), \quad 1$$

where each member of the population is characterized by the duplet (y, z); y denotes the student achievement scores, and z is an indicator specifying whether a student would participate in a survey if sampled.

In IEA studies, the measure of central tendency that is often used to characterize $P(y)$ is its expected value $E[y]$. For each participating population, this parameter is estimated and communicated in the international reports. Given its centrality, in this project we narrowed our attention to the identifiability of $E[y]$.

The Law of Total Expectations tells us that the expected value of y equals the expectation over the conditional expectation of y given a random variable z. That is, $E[y] = E\big[E[y|z]\big]$. This result is useful to study unit nonresponse, as it tells us that the mean achievement in a population is the mean over the conditional mean given participation in the survey. That is,

$$E[y] = E[y|z = 1]P(z = 1) + E[y|z = 0]P(z = 0). \quad 2$$

The inference reported by IEA studies asserts that nonresponse is ignorable within adjustment cells, such that $E[y] = E[y|z = 1]$, since y is mean independent of z by assumption. Section 2 of this report recommends an alternative strategy to report inferences about population parameters. To do so, we explore inferences about $E[y]$ across two different nonresponse models. The first model makes weaker assumptions about $E[y|z = 0]$. This model leads to more credible but also more ambiguous inferences about the expected value of y. The second model is based on the standard nonresponse model maintained in IEA studies. This model makes stronger distributional assumptions, leading to unambiguous but less credible inferences about the parameter of interest. We argue that the reporting strategy we propose would make more explicit the ambiguity in the estimation introduced by nonresponding units in IEA surveys but also inference would be more credible. Moreover, it will provide a useful visualization of the impact that the assumptions maintained in the nonresponse model have on the inference reached. For this, we use data from the ICILS of 2018.

Finally, Section 3 in this report makes recommendations about how to estimate $P(z = 1)$ in Equation 2 for categorizing study populations according to their estimated population participation rate. The minimum participation rate requirements established by IEA standards set out thresholds for $P(z = 1)$ under which point-estimates about $E[y]$ are deemed credible. We recommend that, if an estimate $P(z = 1)$ is used to discriminate between credible or not credible inference when point-estimating $E[y]$ under the maintained nonresponse model. then the estimate of $P(z = 1)$ should not be constructed under the same nonresponse model. In this section, we recommend estimators of $P(z = 1)$, which are derived and justified in-depth on the accompanying manuscript, that use the empirical evidence alone.

2. Reporting interval and point estimates of $\boldsymbol{E}[\boldsymbol{y}]$

In this section, we recommend a complementary form of reporting inferences about $E[y]$. This recommendation is motivated by the idea that a logical starting point when making inferences about a parameter of interest is to determine what can be learned from the data alone, while making minimal distributional assumptions about nonparticipating students (Manski, 1995). We use data from ICILS 2018 to showcase how this can be achieved in an internationally comparative way. The parameter of interest is mean achievement on computer and information literacy (CIL) measured in the ICILS. In this context, achievement is a latent construct defined as the "ability to use computers to investigate, create and communicate in order to participate effectively at home, at school, in the workplace and in society'' (Fraillon et al., 2019, p. 53).

For each study population participating in ICILS, Figure 1 reports inference about $E[y]$ under two alternative set of assumptions about $E[y|z=0]$. First, each vertical line depicts the estimated identification region of $E[y]$ under the assumption that, within each study population, the mean score among nonparticipating students lies within the interval bounded by the estimated fifth and ninety-fifth percentile of the observed distribution of CIL achievement $P(y|z=1)$. That is, $P_{05}(y|z=1) \leq E[y|z=0] \leq P_{95}(y|z=1)$. No other assumption about nonresponding students is imposed in the estimation. This estimation alternative is analogous to $H_{A1}\{\hat{\mu}\}$ in the accompanying manuscript (with $\alpha = 0.05$). Second, for each study population, Figure 1 depicts with a dot the same estimate of $E[y]$ as the one reported by the ICILS International Report (Fraillon et al., 2019, p. 75). That is, this estimate is built upon the assumptions embedded in the standard nonresponse model applied in IEA studies. That is, (1) the mean score among recipient and donor school is the same, and (2) nonresponse is ignorable within adjustment cells after the inclusion of donor schools.

Figure 1 makes explicit that, when abstaining from strong assumptions, inference about $E[y]$ is more ambiguous in study populations with higher nonresponse rates. For example, the nonresponse rate in the USA is about 38% and is the highest across all study populations. The USA also has the widest estimated identification region, with a width of about 102 score points. Accordingly, the USA was labeled as *category 3* in the ICILS International Report. Denmark (DNK), Portugal (PRT), and Uruguay (URY) were labeled as *category 2*, while all other study populations were labeled as *category 1*.

We find it relevant to highlight that, among study populations labeled as *category 1*, IEA standards attach the same credibility to the inference reached by the point estimate. However, Figure 1 makes it clear that, under weaker and more credible assumptions, the ambiguity on the inference among this set of study populations varies substantially, to the extent of that the width of the estimated identification region ranges from 80 score points in Germany (DEU) to nine points Kazakhstan (KAZ). Consequently, the inference reached by the point-estimate in DEU relies more intensively on the nonresponse model maintained by ICILS, than the inference reached in KAZ.

To conclude, Figure 1 reports estimation of $E[y]$ under the two extreme, but not contradictory, alternatives. First, the inference derived from the identification region is the most credible, to the extent that it imposes the least restrictions on $E[y|z=0]$. However, this inference is the most ambiguous, as it leads to identification regions whose length are proportional to $P(z=0)=1-P(z=1)$. On the other hand, inference reported by IEA studies is unambiguous but very restrictive about $E[y|z=0]$ and therefore more questionable. Finally, it is to be noted that for each study population, the point estimate lies within the estimated identification region, which suggests that the alternatives presented do not contradict each other but are complementary from a policy perspective.

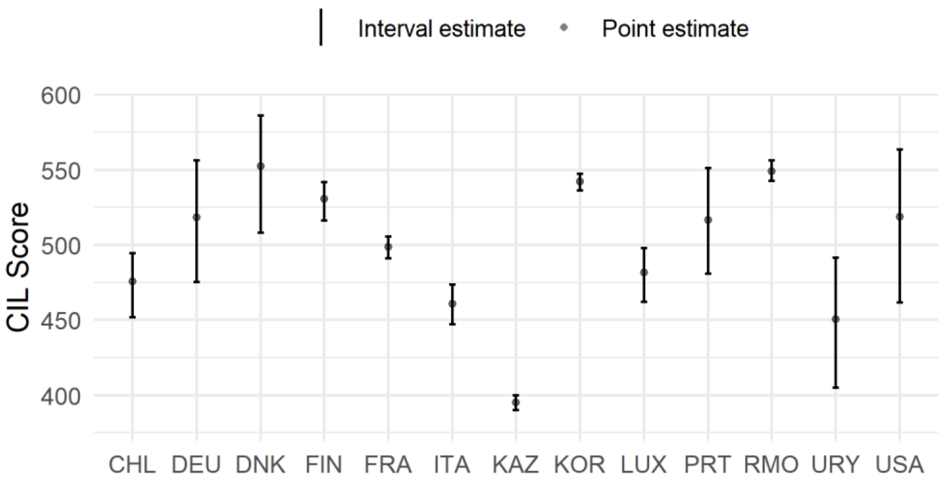

Figure 1 Interval and point estimation of CIL mean score.

__Recommendation 1__: *We recommend reporting in each IEA Study inference about $E[y]$ while making minimal distributional assumptions about nonparticipating students. This can be done either in the International Report if one wants to reflect on the ambiguity in the inference caused by nonresponse, or in the Technical Report if one wants to reflect on the impact of the nonresponse model from a technical perspective. Figure 1 provides an example on how to achieve this.*

3. Estimating $P(z=1)$ for assessing the credibility of point-estimating $E[y]$

IEA studies report estimates of $P(z=1)$ to assess and justify the credibility of the inference reached with point-estimates of $E[y]$ under the mainlined nonresponse model. That is, when $P(z=1)$ is estimated to be below some threshold, IEA studies find it hard to justify the nonresponse model and therefore point-estimates of $E[y]$ are categorized as questionable. In this section, we recommend that, if $P(z=1)$ is used to discriminate between credible or not credible inference about $E[y]$, under the maintained nonresponse model. then the estimate of $P(z=1)$ should not be constructed under the same nonresponse model. As suggested in the accompanying manuscript, the sampling process generating the empirical evidence in IEA studies identifies $P(z = 1)$. Therefore, inference about this population parameter can be reached without assumptions.

<u>**Recommendation 2**</u>: *We recommend estimating $P(z = 1)$ using the empirical evidence alone, if such an estimate is used to assess the credibility of point-estimates of $E[y]$ under the maintained nonresponse model.*

In the accompanying manuscript, it is argued that to point-estimate the population participation rate $P(z = 1)$, under an implemented sampling plan, it is useful to break down estimation into two parts: (1) $\hat{p}_{sch}$ the proportion of students in the population that attend participating schools, and (2) $\hat{p}_{std}$ the population proportion of students participating in the survey within participating schools. While $\hat{p}_{sch}$ is identified by the first-stage sampling process, $\hat{p}_{std}$ is identified by the second-stage sampling process. A point-estimate of $P(z = 1)$ is given by $\hat{p} = \hat{p}_{sch} * \hat{p}_{std}$.

In the manuscript it is proposed estimating the proportion of students in the population that attend participating schools with the following expression:

$$\hat{p}^{1}{}_{sch} = \frac{\sum_{j(part)} MOS_J * (1/\pi_j)}{\sum_j MOS_J * (1/\pi_j)}, \qquad 2$$

where $j = 1, \dots, J$ indexes sampled schools, and MOS_J is the measure of size used to determine the event probability that school j is included into the survey (π_j).

Conversely, IEA studies estimate this same population parameter with the following expression[2]:

$$\hat{p}^{2}{}_{sch} = \frac{\sum_{j(part)}\left[(1/\pi_j) * \sum_{i \in j}(1/\pi_{i|j})\right]}{\sum_{j(part)}\left[(1/\pi_j \pi_s^f) * \sum_{i \in j}(1/\pi_{i|j})\right]}, \qquad 3$$

where $\pi_s^f = \frac{n(part)}{n(samp)}$ is the so-called school-level nonresponse adjustment factor in a stratum (i.e., under the maintained nonresponse model); and the component $\sum_{i \in j}(1/\pi_{i|j})$ reflects the measure of size of a school during data collection.

In this report, we abstract from the discussion of whether MOS_J (as in Equation 2) should be used rather than $\sum_{i \in j}(1/\pi_{i|j})$ (as in Equation 3) as a measure of size for school j. We instead focus attention on the denominator of both expressions. Note that $\sum_j MOS_J * (1/\pi_j)$, in Equation 2, estimates the population size with all available information, exploiting the fact that the randomization principles used to select the school sample identify the parameter of interest. Conversely, $\sum_{j(part)}\left[(1/\pi_j \pi_s^f) * \sum_{i \in j}(1/\pi_{i|j})\right]$, in Equation 3, estimates the population size with partial information

[2] For expositional reasons we deviate slightly from the notation used in the technical reports (see, for example, Tieck (2020, p. 84)).

and the assumption that school-level nonresponse is ignorable. This assumption is asserted in the nonresponse adjustment factor $(1/\pi_s^f)$, which claims independence between school participation and the variables determining the selection probabilities of schools. Not only the assumption is unnecessary in this context, as all information required for inference is available; but more importantly, the assumption can be questionable provided the anecdotal evidence that schools with larger probability of selection tend to decline participation more often.

We understand that during data collection, $\sum_{i\in j}(1/\pi_{i|j})$ reflects more accurately the population size within school j than MOS_J. However, we argue that estimation of $P(z=1)$, as a measure of quality or credibility in the inference, is better justified with an assumption-free estimator, rather than with an estimator that has potentially less sampling variance but relies on strong assumptions.

Recommendation 2a: *We recommend using Equation 2 for estimating the proportion of students in the population that attend participating schools.*

We move now forward to the estimation of the proportion of students participating in the survey within participating schools. The estimator proposed in the accompanying manuscript is virtually identical to the one commonly used by IEA studies, except for the inclusion of replacement schools. IEA studies estimate this population parameter using data collected from originally sampled and donor schools. That is, inference about this parameter relies on the assumption that donor and recipient schools have the same distribution of within-school participation. As above, this assumption is unnecessary and to some extent questionable. It is unnecessary as the set of originally sampled schools identifies the parameter of interest. Moreover, we find it hard to justify that the distribution of within-school participation in donor and recipient school is similar, simply because recipient schools do not participate.

Here, again, we argue that estimation of $P(z=1)$, as a measure of quality or credibility in the inference, is better justified with an assumption-free estimator, rather than with an estimator that has potentially less sampling variance but relies on strong assumptions.

Recommendation 2b: *We recommend estimating the population proportion of students participating in the survey within participating schools using only originally sampled schools.*

References

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: IEA international Computer and Information Literacy Study 2018 International Report*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Manski, C. F. (1995). *Identification problems in the social sciences.* Harvard University Press.

Manski, C. F. (2003). *Partial identification of probabilities distributions: Springer series in statistics.* Springer

Martin, M. O., Rust, K., & Adams, R. J. (1999). *Technical standards for IEA studies*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Tieck, S. (2020). Sampling design and implementation. In Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D (Eds.), *IEA International Computer and Information Literacy Study 2018: Technical report* (pp. 59-78). Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).