PIAAC ROUND 1
NONRESPONSE BIAS ANALYSIS SUMMARY

PREPARED BY WESTAT

Missing data can occur when some of the adults selected in the sample are not contacted or refuse to participate (referred to as unit nonresponse), they fail to respond to a particular survey item (referred to as item nonresponse), or because data collected from the sampled adults is contaminated (and thus not useful) or lost during or after the data collection phase. Under ideal situations, every eligible adult in the target population would have a nonzero chance of selection in a national sample, and would be located, agree to participate in the study, and respond to all survey items. In practice, these circumstances are not realised in any survey population. Nonresponse bias can be substantial when two conditions hold: 1) the response rate is relatively low and 2) the difference between the characteristics of respondents and those of nonrespondents is relatively large. This is reflected in the following deterministic nonresponse bias formula:

$$Bias(\bar{y}_R) = (1 - W_R)(\bar{Y}_R - \bar{Y}_{NR}),$$

where W_R is the proportion of respondents, $\bar{Y}_R$ is the mean outcome for respondents, and $\bar{Y}_{NR}$ is the mean outcome for nonrespondents. An alternative model of nonresponse assumes each sampled person has a certain propensity to respond, and nonresponse bias in a characteristic is a function of the covariance between the response propensity and the characteristic:

$$Bias(\bar{y}_R) = \frac{\sigma_{yp}}{\bar{p}},$$

where σ_{yp} is the covariance between the outcome variable and response propensity, and $\bar{p}$ is the mean response propensity. Based on this model, NRB is present if missingness is related to proficiency, as measured by PIAAC.

Countries worked to reduce nonresponse bias to the extent possible before, during, and after data collection. Before data collection, countries implemented field procedures with the goal of obtaining a high level of cooperation. Most countries followed the PIAAC required sample monitoring activities to reduce bias to the lowest level possible during data collection. Finally countries gathered and used auxiliary data to reduce bias in the outcome statistics through nonresponse adjustment weighting.

All countries were required to conduct a basic nonresponse bias analysis (NRBA) and report the results. The basic analysis was used to evaluate the potential for bias and to select variables for nonresponse adjustment weighting. In addition, countries were required to conduct and report the results of a more

extensive NRBA if the overall response rate was below 70%, or if any stage of data collection (screener, background questionnaire, or the assessment) response rate was below 80%. An item NRBA was required for any BQ item with response rate below 85%.

This report does not include France and Russia as they are on a delayed schedule for submission of the NRBA results.

A summary of the results of the basic NRBA is provided in Section 1. Section 2 contains the results of the extended NRBA, and Section 3 provides a summary of the item nonresponse analysis. A brief summary and conclusions of the NRBA is given in Section 4.

1 BASIC NRBA

The basic NRBA involved comparing survey respondents and nonrespondents using auxiliary variables available on the sampling frame, available from a previous data collection stage (e.g. screener data for the BQ analysis), or coming from an external source that could be matched to each sampled unit. Also, observational data on respondents and nonrespondents collected during data collection could have been used to evaluate bias, assuming the data was of sufficient quality. The auxiliary variables must have been available for all eligible units and, as noted above, had to be related to proficiency. All countries were required to include the following variables in their analysis: age, gender, education, employment, and region. If any of these variables was not available for all eligible units, then a corresponding area-level variable could have been used instead (e.g. the employment rate within small geographic areas).

The basic analysis included results from the following:

- Comparison of response rates for different subgroups;
- Use of a chi-square test or estimates of relative bias to compare the distribution of auxiliary variables (correlated with proficiency) for respondents and nonrespondents; and
- Use of a classification tree algorithm to identify subgroups with low response rates or use of logistic regression to model the relationship between response status and the auxiliary variables.

The response rate and chi-square analyses were useful in explaining the relationship of response status to each auxiliary variable individually. A classification tree algorithm and/or a logistic regression model was used to evaluate the relationship between response status and multiple auxiliary variables.

All countries completed all the required analyses and included all the required variables, age, gender, education, employment, and region, in their analysis, with the exception of Austria, Belgium, Finland, and Italy. In most cases, the failure to include the required variables in the analyses was due to the lack of access to sources with reliable data for such variables.

An initial basic NRBA was conducted prior to the weighting process. The analysis was conducted in two stages. The first stage helped to create a pool of predictor variables related to proficiency, using the field test data. The second stage helped to reduce the pool of predictor variables to those related to response propensity (this was repeated after the weighting process to finalise the basic NRBA). Most countries used all auxiliary variables that showed potential for bias in deriving nonresponse adjustments to the sampling weights. The remaining countries used most of the variables identified in the initial basic NRBA, mainly because reliable data was not available for the remaining variables.

Nonresponse weighting adjustments reduce bias in the outcome statistics to the extent that auxiliary variables are correlated with proficiency. Mainly, weighting adjustments are carried out by assuming nonrespondents' proficiency levels are the same as the respondents in the subgroups created for weighting adjustments using the auxiliary variables. This assumption is, of course, not true and the level of bias

reduction depends on the number of auxiliary variables used during weighting and the correlation between these variables and proficiency.

The basic NRBA is a good initial assessment of nonresponse bias and is essential in identifying effective weighting variables. However, it has its limitations. The analysis does not reflect the effect of weighting adjustments on NRB, and the extent of bias remaining after nonresponse adjustments are conducted. Therefore, countries with lower response rates were required to conduct a more extensive analysis to assess the potential for bias remaining after nonresponse adjustment weighting. Section 2 includes a brief description of the results of the extended NRBA.

2 EXTENDED NRBA

A more extensive NRBA was required if the overall response rate was below 70%, or if any stage of data collection (screener, background questionnaire, or the assessment) response rate was below 80%.

Australia, Korea, and United States achieved an overall response rate of 70% or greater, with response rates for each stage being greater than 80%, and thus did not require the extended NRBA. Cyprus and Ireland also achieved overall response rates of 70% or greater, but they achieved a lower than 80% response rate for one stage of their sample. The remaining countries achieved response rates lower than 70%.

The main purpose of the extended analysis was to assess potential for remaining bias in the final weighted proficiency estimates after adjusting for nonresponse. Because the proficiency levels of nonrespondents are unknown, the NRBA is carried out by making assumptions about nonrespondents. Therefore, it is necessary to conduct multiple analyses to assess the potential for bias since each analysis has its own limitations resulting from the specific assumptions made about nonrespondents. The extended NRBA included seven analyses (as listed below). Together, they were used to assess the patterns and potential for bias in each country data.

The extended NRBA included the following analysis:

1. Comparison of estimates before and after weighting adjustments;
2. Comparison of weighted estimates to external totals;
3. Correlations of auxiliary variables and proficiency estimates;
4. Comparison of estimates from alternative weighting adjustments;
5. Analysis of variables collected during data collection;
6. Level-of-effort analysis; and
7. Calculation of the range of potential bias.

These analyses are described further below.

Cyprus and Ireland were required to do only a subset of the analysis since their overall response rate was higher than 70%.

Comparison of Estimates Before and After Weighting Adjustments

To better capture the effects of the weighting adjustments on unit nonresponse bias, estimates from the full sample were compared to estimates from the respondents before and after weighting adjustments. To compare estimates before and after each step of weighting adjustments, the following comparisons were made:

- Comparison of percentage distributions from BQ base weights for the total eligible sample of persons with the BQ base weights for the BQ respondents to check for differences due to nonresponse to the BQ
- Comparison of percentage distributions from BQ base weights for the total eligible sample of persons with that from the BQ nonresponse adjusted weights for respondents to check for differences after the nonresponse adjustment process to the BQ
- Comparison of percentage distributions from BQ nonresponse adjusted weights for respondents with that from the BQ raked weights (weights adjusted to two or more marginal population totals) for respondents to check for differences that may have been introduced through the initial raking procedure.

For countries that had screeners, analogous comparisons to the BQ level, as mentioned above, were completed. All the countries required to do the analysis completed it. The goal was to include at least one auxiliary variable not present in weighting adjustments in addition to those used during nonresponse adjustment weighting. Inclusion of the non-weighting variables shows whether the weighting adjustment was effective in reducing bias in other known auxiliary variables, not just the weighting variables. The following 11 countries; Denmark, Finland, Germany, Japan, Netherlands, Norway, Poland, Slovak Republic, Sweden, England, and Northern Ireland, included non-weighting variables in this analysis as well as weighting variables. The remaining countries only included the weighting variables. Canada included a substantial number of weighting variables in their analysis. In general, all countries observed that bias was reduced in auxiliary variables through weighting adjustments.

Comparison of Weighted Estimates to External Totals

The second analysis compared estimates from PIAAC to external source estimates to assess potential for bias in PIAAC outcome statistics.

To the extent possible, countries used estimates from external sources that measured the same characteristic for a similar time period. Some external source estimates were subject to sampling error also, and thus the variance of these estimates were taken into account when making comparisons.

Many countries found significant differences between the PIAAC estimates and the external source estimates but were mostly able to explain the sources for discrepancies. The sources mainly included, different data collection time periods or different definitions (e.g., definition of employment). All countries, except the UK completed this analysis.

Correlations of Auxiliary Variables and Proficiency Estimates

The analyses described thus far relied on auxiliary variables and did not directly measure bias in the proficiency estimates. Bias in the auxiliary variables is indicative of bias in the proficiency estimates to the extent that the auxiliary variables and proficiency estimates are correlated. Thus, correlations between the auxiliary variables and proficiency data are good indicators of potential for bias reduction through weighting adjustments. For variables used in the weighting adjustments, a low correlation with proficiency implies that using the variable in the weighting adjustments did little to reduce nonresponse bias. On the other hand, a high correlation with proficiency implies a potentially high reduction in nonresponse bias. However, it should be noted that the correlations are based on respondents' data, and the relationship between proficiency and the auxiliary variables might be different for nonrespondents. Therefore, the correlations could be different if a country's response rate is very low, and if nonrespondents are different from respondents in terms of the relationship between their scores and the auxiliary variables.

Correlations were calculated as the square root of R-square of a weighted analysis of variance, whose dependent variable was the literacy or numeracy score while the explanatory variables were the weighting variables (BQ nonresponse adjustment cells and raking dimensions).

Table 1 presents the correlation between the proficiency and the weighting variables for each country. There are a few countries with low correlation between the BQ nonresponse cells and the proficiency scores. However, all of the correlations between proficiency scores and the BQ nonresponse cells and the raking dimensions combined are higher than 0.35 and the average is 0.52 for literacy scores and 0.53 for numeracy scores. Although it was not required, the correlations for Korea and the US were also provided. Based on the moderate-to-high correlations between the weighting variables and the proficiency scores, we can expect the weighting adjustment to have reduced bias in the proficiency scores.

Table 1. Correlations of Auxiliary Variables and Proficiency Estimates

Country	Literacy	Numeracy
AUT	0.56	0.57
BEL	0.36	0.36
CAN	0.54	0.53
CYP	0.39	0.47
CZE	0.56	0.60
DEU	0.61	0.62
DNK	0.50	0.46
ENG*	0.52	0.56
ESP	0.62	0.62
EST	0.37	0.35
FIN	0.60	0.58
IRL	0.52	0.53
ITA	0.49	0.53
JPN	0.53	0.52
KOR	0.55	0.55
NIR*	0.57	0.60
NLD	0.57	0.55
NOR**	0.45	0.45
POL	0.40	0.37
SVK	0.38	0.38
SWE	0.70	0.70
USA	0.63	0.66

*England and Northern Ireland were weighted separately to allow efficient estimates for each population.

** Norway was not able to provide nonresponse adjustment cells due to confidentiality concerns. The correlation with raking dimensions was 0.23 and 0.22 for literacy and numeracy, respectively. Norway self-reported the correlation between literacy scores and BQ nonresponse adjustment variables as 0.45 and did not report the correlation with both the nonresponse adjustment variables and the raking dimensions. Therefore, 0.45, being the larger of the two correlations, was used to represent the correlation with all weighting variables.

Figure 1 displays each country's correlation between weighting variables and the literacy score and correlation between weighting variables and the numeracy score. The two correlations are very close to each other, implying the same level of effectiveness in reducing bias for the two proficiency estimates.

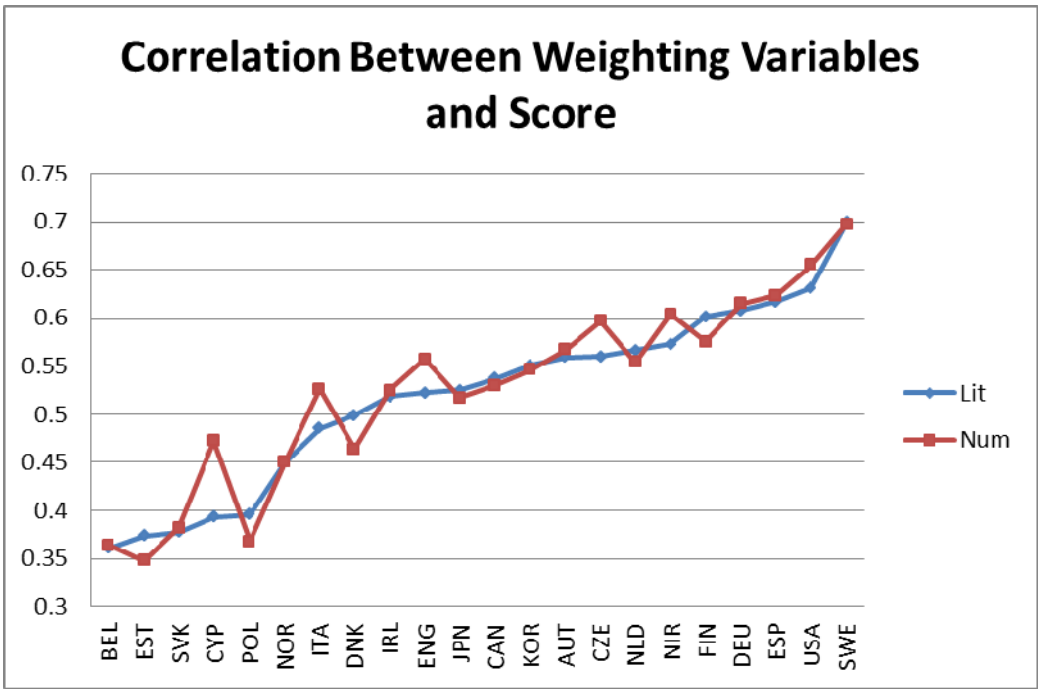

Figure 1. Correlation between weighting variables and the proficiency scores

Figure 2 shows the plot of response rate versus correlation between the weighting variables and the literacy score reflecting the effectiveness of nonresponse adjustments in reducing bias.

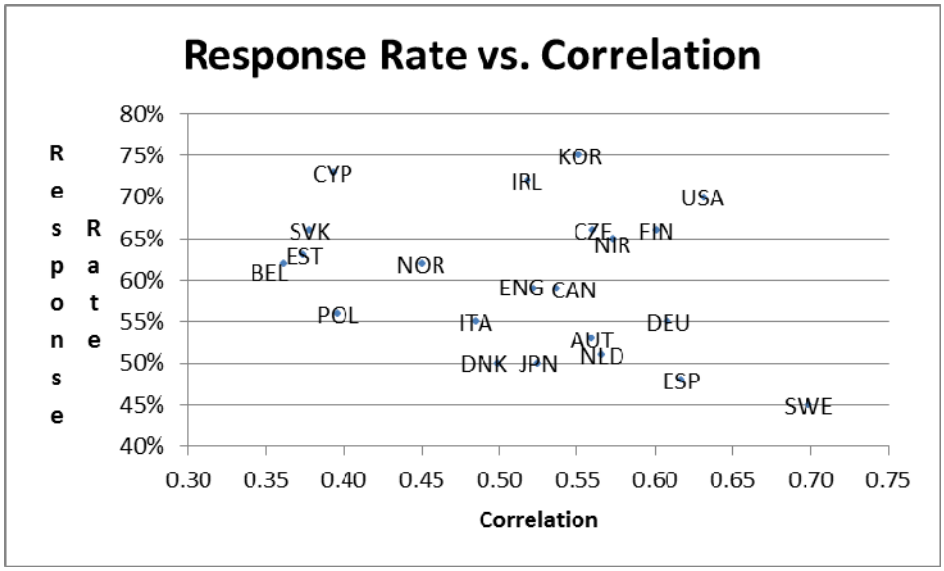

Figure 2. Scatterplot of response rate versus correlation

Figure 2 shows that:

- Countries in the lower right corner, such as Sweden, Spain and Germany, have low response rates, but are expected to have accomplished a considerable bias reduction through weighting, since their weighting variables are highly correlated with the proficiency.
- Austria, Canada, Denmark, England, Italy, Japan, and Netherlands have about average correlations, so bias reduction is expected at an average level as compared to other countries.

- Finland and US have a higher than average correlation and high response rates.
- Belgium, Cyprus, Estonia, and Slovak Republic have low correlations, but relatively high response rates, which helped reduce potential for bias. Poland, which also has a low correlation, has a somewhat lower response rate, which indicates relatively less potential for bias reduction.

Comparison of Estimates from Alternative Weighting Adjustments

For this evaluation, an auxiliary variable was re-calibrated to known totals, and estimates of the key statistics were compared before and after the re-weighting. Re-weighting was useful as an evaluation tool when:

- The variable was not used in weighting (because it was not available) or was used but with different categories;
- The variable is correlated with the outcome measure; and
- The variable is correlated with response propensity.

Any differences between estimates using the official survey weights and the re-weighted weights reflected noncoverage as well as nonresponse bias, but if there was not a large change in the estimates, this was further confirmation that nonresponse bias may not be a concern.

Thirteen of the countries fully complied with the analysis and results confirmed that nonresponse bias may not be a concern. These countries were: Austria, Belgium, Canada, Denmark, Estonia, Finland, Germany, Japan, Netherlands, Norway, Poland, Spain, and Sweden. Italy found a significant difference between the average literacy score using final weights and when using the alternative weights, where the alternative weights were created using a more detailed weighting variable. Some caution should be used in conclusions from this analysis for Czech Republic (quality unknown due to unavailability of data), Slovak Republic (partial compliance), and UK (did not comply).

Japan and Sweden used the results of this analysis to improve their final survey weights.

Analysis of Variables Collected During Data Collection

Disposition codes contain information on reasons for nonresponse. For this analysis, distributions of sampled persons with known characteristics related to outcome (i.e. the literacy-related nonrespondent (LRNR) cases, which are language problems, reading and writing difficulty, and mental disability) were examined. For example, the demographic distribution of literacy-related cases were compared to other eligible persons using auxiliary data, and interview data. Statistical tests such as Chi-square tests were processed to determine if there is a relationship between select demographic variables and the disposition codes for nonrespondents. A special weighting adjustment for literacy-related cases was conducted for all countries, with the exception of Poland, where the BQ LRNRs together with the other BQ NRs were represented by BQ respondents. Therefore, in almost all countries, the existence of LRNR cases were dealt with appropriately in order to reduce potential for bias.

All countries conducted an analysis of disposition codes with some observing differences that were expected, given the conditions in their countries. However, Sweden and the UK each conducted only a partially completed analysis (i.e., the quality level is unknown) due to unavailability of data.

In addition, Non-Interview Report (NIR) forms identify observable demographic information and reasons for nonresponse that are not captured in the disposition codes. The NIR forms can potentially indicate whether the reasons for nonresponse are related to proficiency estimates and suggest ways to improve response rates for future surveys.

The following countries put extra effort in conducting the analysis using the information from NIR forms: Cyprus, Germany, Italy, Japan, and Slovak Republic. The observed information from NIR forms may be useful for data collection in the next cycle.

Level-of-Effort Analysis

Another way to evaluate bias in the proficiency estimates is to compare proficiency estimates by level of effort. To the extent that the late or hard-to-reach respondents are similar to the nonrespondents, differences in proficiency estimates between the late and early (or hard-to-reach and easy-to-reach) respondents could indicate nonresponse bias. This analysis can be useful in detecting potential for bias given the assumption that nonrespondents are similar to respondents at the end of the data collection period.

If the literacy estimates differed between easy and hard respondents within a category of a weighting variable (used in the level-of-effort analysis), that may indicate that there are differences even within the weighting cells, and the nonresponse adjustment might not have helped. However, it may be that the data collection procedures were effective in obtaining a different type of respondent, potentially reducing the bias.

Thirteen countries revealed some significant differences in characteristics between early and late respondents, including Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Italy, Japan, Netherlands, Norway, Poland, Spain and Sweden. Two countries, Finland and Germany, conducted the analysis but did not find significant differences. Slovak Republic did not comply with the analysis, and some caution should be used in drawing conclusions from UK's analysis due to unavailability of data.

Calculation of the Range of Potential Bias

The final component of the bias analysis is to evaluate the potential for bias remaining after weighting under the scenario that nonrespondents' proficiency scores are vastly different from the assumptions made during weighting.

It is well known that NRB can be reduced to some unknown extent through sample weighting when proficiency is correlated with auxiliary variables, and auxiliary variables are correlated with response propensity. Weighting assumes response probabilities are constant within every group created for weight adjustment, the proficiency score has zero variance within each group, and response propensity is uncorrelated with proficiency. It is known that these assumptions are not correct, and the impact of weight adjustments is limited to the number of variables available for nonresponse adjustment, and correlation levels with proficiency. Also, it is not possible to measure the exact departure from these assumptions since proficiency levels of nonrespondents are not known. This analysis attempts to evaluate the potential for bias by computing a range based on an extreme assumption that nonrespondents would all score at the 10^{th} percentile within each weighting cell, and at the other extreme they would all score at the 90^{th} percentile within each weighting cell. The range of bias was computed as the difference between the two extreme estimates, while taking into account the response rate and population size in the weighting cell.

The literacy scores' first plausible value was used to compute the range of scores within the responding sample and to predict the range of estimates for nonrespondents.

If the weighting classes were well defined, that is, each weighting class successfully contains a homogeneous population in terms of proficiency scores, then scores would not vary much within a weighting cell, so the range of bias would be small. On the other hand, the range of bias is also affected by the response rate. If the response rate is high, the range of bias may not be high even when the respondents have a wide range of scores in the weighting cell, because the proportion of nonrespondents whose score will get filled in with the extreme values is low. Thus, the range of bias analysis measures

the impact of response rate on the quality of final estimates as well as the effectiveness of the weighting adjustments in reducing the potential for bias.

Figure 3 displays the range of potential bias in outcome statistics after weighting adjustments are incorporated in the official weights. For comparison purposes, the range of bias before weighting is included in the figure also. The range of bias before weighting was computed without regard to weighting cells, based on the extreme assumption that nonrespondents would all score at the 10th percentile, and at the other extreme they would all score at the 90th percentile. The countries are sorted by their nonresponse rate and each country's nonresponse rate is shown in Figure 4. Figure 3 shows that the range of bias after weighting adjustment is significantly lower than before weighting adjustments are conducted, that is, each country data achieved a substantial bias reduction through nonresponse adjustment weighting. In addition, countries with higher response rates, such as Ireland, Slovak Republic, and Estonia, have lower range of bias. However, some countries with a low response rate, such as Sweden, Spain, Denmark, and Japan, have low ranges of bias also, due to their effective nonresponse adjustment weighting processes. Results of from the UK were inconclusive due to partial compliance. The results from the range of bias analysis re-emphasises the importance of minimising bias in the sample throughout the survey process, and achieving high response rates especially if the country does not have access to auxiliary variables highly correlated with proficiency.

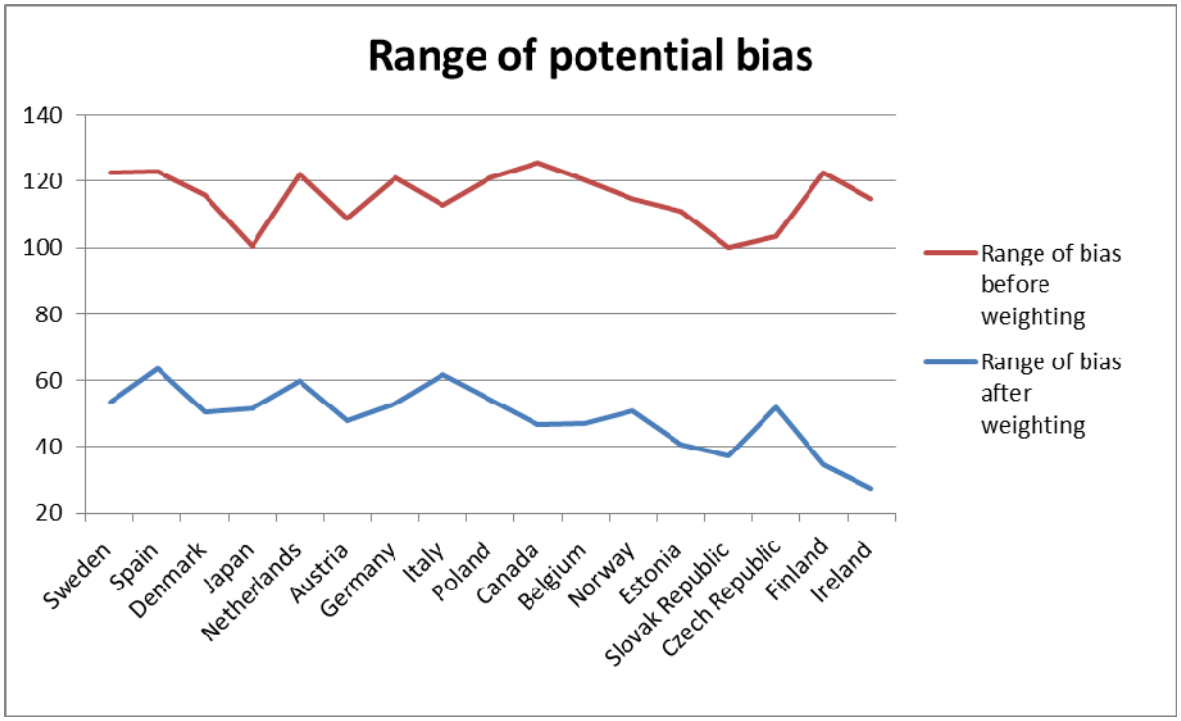

Figure 3. Range of potential bias before and after weighting adjustment

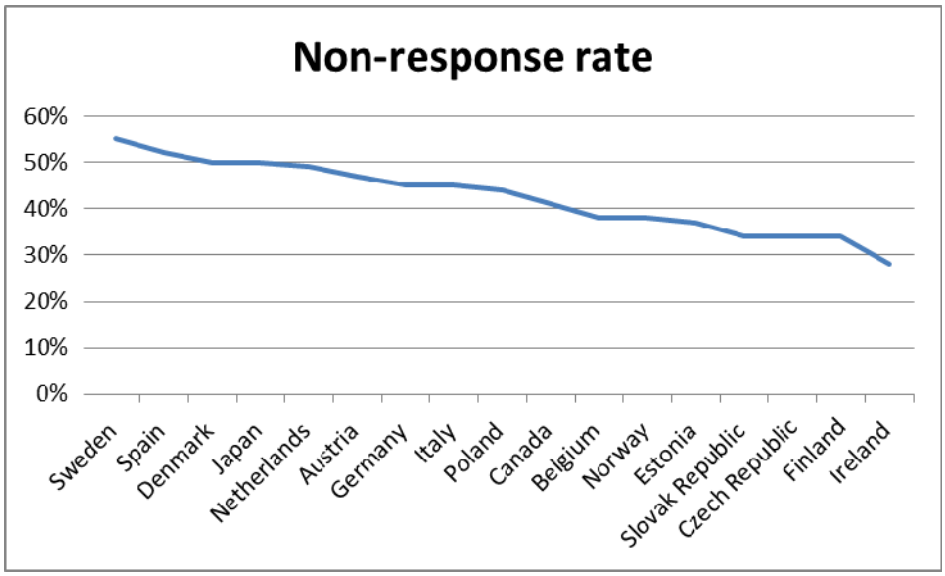

Figure 4. Nonresponse rate by participating country

3 ITEM NRBA

Countries were required to conduct nonresponse bias analysis for any BQ item with a response rate below 85%. Only two items showed low response rates: item D_Q17B (Earnings – additional payment amount last year), and item D_Q18A (Earnings – total earning last year).

Czech Republic, Estonia, Italy, Poland, and Slovak Republic were the only countries that had less than an 85% response rate for either D_Q17B or D_Q18A, with the lowest response rate being equal to 75% for D_Q18A for Poland.

4 SUMMARY AND CONCLUSIONS

PIAAC standards were established with the main goal of producing reliable and comparable data across participating countries. As a result, a number of standards and guidelines were developed to help countries achieve the highest response rate possible, and at the same time reduce nonresponse bias to the minimum achievable. In addition, all countries were required to conduct a basic NRBA, and countries with lower response rates were required to conduct an extended NRBA.

All countries were required to conduct a basic nonresponse bias analysis (NRBA) and report the results. In addition, countries were required to conduct and report the results of a more extensive NRBA if the overall response rate was below 70%, or if any stage of data collection (screener, background questionnaire, or the assessment) response rate was below 80%. An item NRBA was required for any BQ item with response rate below 85%.

The basic and extended NRBA included several analyses. Each analysis was based a number of assumptions about nonrespondents, limiting the utility of the results. Thus, multiple analyses were used to assess the potential for bias in outcome statistics.

Correlation between the auxiliary variables used during weighting and the proficiency scores is a good indication of the effectiveness of nonresponse adjustment weighting. A number of countries with low response rates had higher correlations, implying a more effective nonresponse adjustment than countries with lower correlations. However, data users need to be cautioned that the analysis is based on

correlations between respondents' proficiency scores and the auxiliary variables. That is, the analysis assumes that the same correlations exist for the remaining sampled cases that have no scores.

Table 2 summarises the results of the NRBA for countries with response rates lower than 70%. The analysis showed that nonresponse adjustment weighting was effective in reducing the potential for bias in all countries. Countries that achieved higher response rates guaranteed a minimised level of bias in outcome statistics, whereas countries with lower response rates had to rely on the auxiliary variables available to them for nonresponse adjustment. Countries with relatively higher response rates and highly effective nonresponse adjustment showed minimal potential for bias as compared to countries with lower response rates, or those with moderately effective nonresponse adjustment weighting. The results for the UK are inconclusive because many of the analyses were either incomplete or not conducted.

The analysis concluded that there was not enough evidence showing any moderate or high level of bias in the outcome statistics across the countries. However, this conclusion was based on assumptions made about the proficiency scores of nonrespondents. Therefore, data users need to be cautioned when interpreting the results of the NRBA for countries with very low response rates since different assumptions could lead to different results. For example, a response rate of 50% would mean making assumptions about half of the sample with no data. Multiple analyses, with different assumptions, were included in the NRBA to protect against misleading results, however, the lower the response rate, the higher is the risk of hidden biases that are undetectable through NRBA even when multiple analyses are involved.

Table 2. PIAAC NRBA Outcome Summary for Countries with Response Rates Lower Than 70%

Country	Outcome
Austria (AUT)	Caution-Bias low
Belgium (BEL)	Caution-Bias low
Canada (CAN)	Caution-Bias minimal
Czech Republic (CZE)	Caution-Bias low
Denmark (DNK)	Caution-Bias low
Estonia (EST)	Caution-Bias low
Finland (FIN)	Caution-Bias minimal
Germany (DEU)	Caution-Bias low
Italy (ITA)	Caution-Bias low
Japan (JPN)	Caution-Bias low
Netherlands (NLD)	Caution-Bias low
Norway (NOR)	Caution-Bias low
Poland (POL)	Caution-Bias low
Slovak Republic (SVK)	Caution-Bias low
Spain (ESP)	Caution-Bias low
Sweden (SWE)	Caution-Bias low
UK – England	Caution-Bias Unknown
UK – Northern Ireland	Caution-Bias Unknown