

Corresponding author: John Micklewright

Affiliation: Institute of Education, University of London

Email address: j.micklewright@ioe.ac.uk <<mailto:j.micklewright@ioe.ac.uk>>

Co-authors: Sylke Schnepf and Chris Skinner (both of the University of Southampton)

Title: Non-response biases in the PISA survey of school children: an assessment for England by combining survey data with administrative registers

Judgment of whether non-response affects survey data quality requires the pattern of response to be assessed and not just the level. Researchers typically struggle to make an adequate assessment of response patterns, with analysis often limited to variables available on the survey's sampling frame. These variables may only be weakly correlated with response and with outcome variables of interest. This is the situation prevailing in existing analyses of non-response biases in international surveys of children's learning, including in the PISA survey conducted in OECD countries that we focus on in this paper. By contrast, we assess response biases with rich auxiliary information on cognitive achievement from administrative registers that is available for all sampled units in our survey data. This information is highly correlated with the main survey outcome variables - scores in tests of ability in reading, maths and science. By combining our survey data with data from administrative registers in this way we are able to undertake an unusually rich analysis of the pattern of survey non-response and of the resulting biases. We consider biases in survey estimates of mean achievement, variance of achievement, the percentage of children below a given achievement threshold, and in regression parameters linking achievement to family socio-economic status. The analysis is carried out for England, profiting from the existence in this country of national tests of learning and public exams that are taken close to the age of the PISA target population (15 years) and which are recorded in administrative registers to which we are able to link the PISA sample. Interest in our analysis is strengthened by the exclusion of the UK by the OECD from published PISA results for 2003 on account of fears of the impact of non-response on data quality in England. We analyse biases both for 2003 and for the previous survey round in 2000 when the quality of the data in England was also questioned but without the subsequent exclusion of the UK from published results. PISA has a two stage sample design, with schools being sampled first and then children within schools. We are able to assess biases arising from non-response at each of these stages. Results show that relative to the impact of sampling variation, the biases in estimates of all parameters that we consider are large. In terms of country rankings however, England would move only a small number of places. Estimates of the biases are very similar for 2000 and 2003, calling into question the OECD's practice of using levels of response as the basis for inclusion of a country in the published reports - England met the thresholds in 2000 but not in 2003.