I Z A

IZA DP No. 8411

Which Schools and Pupils Respond to Educational Achievement Surveys? A Focus on the English PISA Sample

Sylke V. Schnepf
Gabriele B. Durrant
John Micklewright

August 2014

Which Schools and Pupils Respond to Educational Achievement Surveys? A Focus on the English PISA Sample

Sylke V. Schnepf
European Commission, DG Joint Research Centre
and IZA

Gabriele B. Durrant
University of Southampton

John Micklewright
IOE, University of London
and IZA

ABSTRACT

Which Schools and Pupils Respond to Educational Achievement Surveys? A Focus on the English PISA Sample

Using logistic and multilevel logistic modelling we examine non-response at the school and pupil level to the important educational achievement survey Programme for International Student Assessment (PISA) for England. The analysis exploits unusually rich auxiliary information on all schools and pupils sampled for PISA whether responding or not, including data from two large-scale administrative sources on pupils' results in national public exams, which correlate highly with the PISA target variable. Results show that characteristics associated with non-response differ between the school and pupil levels. The findings have important implications for the survey design of education data.

Corresponding author:

Sylke V. Schnepf
European Commission
DG Joint Research Centre
Econometrics and Applied Statistics Unit
Via E. Fermi 2749, TP 361
Ispra (VA), I-21027
Italy
E-mail: Sylke.Schnepf@jrc.ec.europa.eu

1. Introduction

Results of international educational achievement surveys make it possible to compare countries' success in educational outcomes and have therefore impacted widely on education policy debate in most of the countries covered by these surveys. However, attention to the survey results has not been matched by thorough reviews of the data quality of such surveys, in particular with respect to non-response rates and non-response bias. This stands in contrast to household surveys, for which considerable research on the nature and correlates of non-response is available (for a summary see Groves (2006)).

The most prominent achievement surveys are the Programme for International Student Assessment (PISA) (Organisation for Economic Co-operation and Development (2012)), the Trends in International Maths and Science Study (TIMSS) (Olson *et al.* (2008)), and the Progress in Reading Literacy Study (PIRLS). The typical design involves sampling schools at a first stage and then pupils within schools at a second stage. The success of the survey depends on the level and pattern of response at both stages. Although a household survey may often have a multi-stage element, with individuals within each contacted household choosing whether to respond, the two stages in a survey of school children are somewhat distinct in a way that is fundamental to the survey design. Non-contact at school or pupil level is likely to be negligible for achievement surveys while factors impacting on school response may be related to the environment of the school setting as well as school commitments. Pupils' decisions to respond may depend on socio-demographic characteristics as well as the level of their schools' commitment. For these reasons, response to this type of survey deserves more attention from researchers than it has received to date.

Non-response analysis often suffers from a lack of information on the non-responding units, in particular on variables correlated with survey target variables. By contrast, our analysis exploits rich auxiliary information on all schools and pupils sampled for PISA using two external administrative datasources: the Pupil Level Annual School Census and the National Pupil Database. That is, we have additional information on schools and pupils irrespective of whether they respond to the survey. This is a rare example of linked survey and administrative data (see also for example Dearden *et al* (2011)). In particular, we have access to administrative registers that contain information on results in national public exams. These are strongly correlated with the main PISA target variables that measure learning achievement – a characteristic highly desirable for non-response investigations (Little and Vartivarian (2005)), but rarely found in practice (Kreuter *et al* (2010)). Even in the presence of linked administrative data such information is often only available for the responding cases. Moreover, in the case of the

schools, we have information on the exam results of all their pupils of the target age for PISA (age 15) and not just on the sample drawn for the survey. References that analyse non-response in education data are limited but include Steinhauer (2014), Rust *et al.* (2013), Pike (2008), Porter and Whitcomb (2005).

This paper models non-response at both the school and pupil stages to PISA in England in the first two rounds of the survey, held in 2000 and 2003. School response among schools initially sampled averaged 66% in these two rounds and pupil response 77%, well below the targets for schools and pupils of 85% and 80% respectively set by the PISA organisers (Our figures for school and pupil response differ slightly from the official published figures for reasons described in Section 2). In fact, international reports for the 2003 round of PISA excluded the UK following concerns that the quality of data for England, where there is a separate survey, suffers from non-response bias due to the lower response rates (Organisation for Economic Co-operation and Development (2004)). PISA's main aim is to assess pupils' learning achievement. Hence, a particular focus of the paper is on the relationship of response at both the school and pupil stages of the survey with the level of pupil ability – mean pupil ability in the school in the case of school response and the individual's own ability for pupil response. These relationships determine the extent of non-response biases in the estimates of learning achievement that the survey provides. A companion paper estimates the size of these biases using different weighting methods, a subject we do not consider here (Micklewright *et al.* (2012)). In the present paper, we provide much more detailed modelling of both school and pupil non-response. At the pupil stage we allow for the influence of the school on pupils' response behaviour. Using cross-level interactions between pupil and school characteristics we can, for example, examine whether high performing students in a low performing school are less likely to respond than high performing students in a high performing school. Furthermore, we use random slope models to assess the school influence on the most important student response predictor: student ability.

This paper is structured as follows. The PISA survey and the linked administrative data are described in Section 2. Section 3 presents response probability models for both schools and pupils. At the school level, we estimate simple logistic regression models. To analyse pupil response we estimate multilevel models that allow investigation of the context-dependent response behaviour of pupils within schools. The results for both models are discussed in Section 4. A summary of the main findings and their implications for educational policy and survey practice are provided in the final section.

2. The PISA survey and the administrative education data

2.1. PISA sample design

The analysis in this paper uses data from the first two rounds of the PISA study in England. The following describes the PISA sampling design. Although small changes in the design have occurred since then - and these are indicated - we do not believe that these changes impact on the interpretation and implications of our results. The PISA survey has a two stage design. First, schools with 15 year olds are sampled with probability proportional to size (PPS). Second, a simple random sample of 35 pupils aged 15 is drawn for each responding school (since 2009 30 students are selected). If a school has fewer than 35 pupils of this age, all are included in the second stage sample. Pupil sampling is carried out from lists provided by schools. The first stage sampling is stratified on school size and type of school – state, quasi-state, and private (the English terms for these three types are 'LEA', 'grant maintained' and 'independent'). Within the first two types, further strata are created on the basis of region and, importantly, the age 16 public exam records that form part of our auxiliary information. Private schools are stratified by region and pupils' gender.

As is common with the international surveys on children's learning achievement, a system of 'replacement' of non-responding schools is operated. Replacement in survey sampling is the subject of considerable debate (e.g. Vehovar (1999); Prais (2003); Adams (2003); Lynn (2004); Sturgis *et al.* (2006)). The PPS sampling generates a list of 'initial schools' together with two potential replacements, the schools that come immediately before and after within the stratum. If an initial school declines to participate, its first replacement is approached. If this first replacement does not respond, the second replacement is asked to participate. Schools sampled in England, including the replacement schools, numbered 543 in 2000 and 570 in 2003 – 181 schools in each group in the first year and 190 in the second.

[Table 1 about here]

Table 1 shows that response rates for initial schools are low with just 60% in 2000 and slightly better with 73% in 2003. However, the majority of replacement schools did not respond, i.e. for both years merged only 41% of first and 29% of second replacement schools responded. These schools faced a higher survey burden. They first had less time to conduct the survey and second the PISA survey data collection was coming to closer proximity to a nationally organised test pupils sit (GCSE exams), the results of which are used for school league tables in England. (Since 2006, the England PISA test takes place after the GCSE exams, so that the survey burden for replacement schools is considerably decreased.) Given this design Sturgis *et al.* (2006)

4

questions whether replacement schools indeed improve non-response bias or if they might actually increase it. In fact, they conclude that this is an untestable question, while this paper will be able to examine this.

Table 1 and the following data analysis are based on a slightly changed school set compared to that of the PISA sample. PISA organisers exclude schools in which less than 25% of students responded (3 schools) which we include. (Nowadays schools with less than 50% student response are excluded. For England across all years until 2012 this results in 7 schools being deleted from the sample.) Given that we merge 2000 and 2003 data in order to increase the sample size, we exclude 42 schools in 2003 that had also been approached in 2000 in order to avoid double counting and dependence of error terms (80% of these schools showed the same response behaviour in both years).

At the second stage, pupils are sampled within schools without replacement. Parents' of sampled pupils receive a letter that describes the survey and explains that children's participation is voluntary. Non-responding students are therefore those who did not attend the class during the time the PISA test was sat either due to parents' lack of consent or other reasons like illness. Table 2 shows that 77% of students respond in our merged sample for 2000 and 2003.

[Table 2 about here]

2.2. Auxiliary information from administrative data

The main aim of PISA is to measure pupils' achievement. If non-response is associated with ability then the survey outcome will be upwards or downwards biased. It follows, that any kind of response modelling to the survey ideally needs to build on good auxiliary information on children's ability. If the auxiliary variables are highly correlated with the survey outcome variable and we do not find an association between our auxiliary variable and response then we can assume that the final sample provides a good estimate of the target variable. England is unusual in having nationally organised exams for the population of children at several ages. Once linked to the PISA survey data, this information allows respondents and non-respondents to be compared on the basis of assessments taken shortly after the survey was conducted. Here we exploit data from the Pupil Level Annual School Census and the National Pupil Database. In particular, we use information from 'Key Stage 4' (KS4) public exams taken at age 15 or, more commonly, at 16. We also experimented with variables derived from 'Key Stage 3' (KS3) national tests in English, maths and science taken typically at age 14 (and compulsory in state schools in both 2000 and 2003). However, conditional on KS4 results these variables did not display any additional explanatory power. Key Stage 4 refers to General Certificate of Secondary Education

exams taken in a wide range of subjects and General National Vocational Qualifications, known respectively as GCSEs and GNVQs. After experimenting with different KS4 variables that are highly correlated, we settle with the Key Stage 4 point score (KS4 ps) for the best 8 subjects, a variable often used in education research in the UK.

[Figure 1 about here]

Figure 1 shows that our auxiliary achievement measure is highly correlated with the PISA test score for reading for responding pupils in 2000 and 2003 (r=0.81). The correlation coefficient of our auxiliary variable with the PISA math test score is 0.79 (based on 2,181 responding students) and with PISA science is 0.80 (based on 2,177 students) – results not shown. This analysis therefore greatly benefits from access to highly associated and fully observed auxiliary information.

Further auxiliary information is also available from administrative records on the child's gender and whether he or she receives Free School Meals (FSM), a state benefit for low income families, and we use this information in Section 4. Information on FSM is not available at the pupil level for 2000 although we do know the percentage of individuals receiving the benefit in the child's school.

We have access to the auxiliary information just described for (almost) all 15 year olds. While we have information on the number of pupils attending a school and whether it is private or public, we derive in addition the following school variables by averaging the PISA target population on this variable within schools: mean KS4 score, percentage of students with free school meal and percentage of male students within a school. This information is contained in the Pupil Level Annual School Census and the National Pupil Database, a combination we refer to as the 'national registers'. The linked data set of PISA survey and national register data was created by us using files supplied by the Office for National Statistics (ONS), as the former survey agency for PISA and the Fischer Family Trust, who extracted data for the population from the national registers. The linking used unique school and pupil identifiers that ONS and the Trust had added to the files. Further details of the files and linkage are given in Micklewright *et al.* (2012).

3.   Methodology

A sequential modelling approach is employed. First response at the school level using a (single-level) logistic model is used analysing the effects of school characteristics on the school response propensities. Then for responding schools, pupil level response within schools is modelled using a multilevel logistic model taking account of both individual- and school-level effects. The

sequential model is a commonly used method in the non-response modelling literature, for example when modelling first contact and subsequent cooperation (Hawkes and Plewis (2006); Lepkowski and Couper (2002)) although sometimes only one of the two processes may be modelled (Pickery *et al.* (2001); Durrant and Steele (2009); Durrant *et al.* (2011)).

Let y_j denote the binary response indicator for school j , $j = 1,...,J$, coded

$$y_j = \begin{cases} 1 & \text{school response} \\ 0 & \text{school non-response.} \end{cases}$$

The response probabilities are denoted by $\pi_j = \Pr(y_j = 1)$. The equation of the logistic regression model (Agresti 2013) is

$$\text{logit}(\pi_j) = \log\left(\frac{\pi_j}{1 - \pi_j}\right) = \alpha^T z_j, \tag{1}$$

where z_j is a vector of school-level covariates and α is a vector of coefficients. To ease interpretation, the following equation is used for judging the importance of the coefficient a_k on the kth explanatory variable z_k (Gelman and Hill (2006)):

$$\frac{d\pi}{dz_k} = \pi(1 - \pi)a_k. \tag{2}$$

To investigate non-response at the pupil level a multilevel modelling approach is employed, recognising the clustering of pupils within schools. Multilevel modelling has become very popular in the education literature to analyse pupil and school effects on educational attainment. Just to name a few, examples include Leckie and Goldstein (2011), Rasbash *et al.* (2011), Blatchford *et al.* (2002) and Gray *et al.* (2001). Multilevel models have also been increasingly used for response analysis, for example accounting for the clustering of individuals or households within interviewers and areas (Pickery *et al.* (2001); Durrant and Steele (2009); Durrant *et al.* (2010)). So far not many researchers have analysed non-response in education surveys using multilevel modelling. For an example see Steinhauer (2014). Besides Micklewright *et al.* (2012) we are not aware of any literature discussing non-response in cross-national education surveys.

For the application here, the multilevel approach has a number of attractions both methodologically and substantively. First, estimates of standard errors account for the clustering of pupils within schools in PISA, thus recognising that pupils cannot be treated as independent draws from the national pupil population. Failure to account for the clustering by school leads to downward bias in standard errors, which in turn leads to overstatement of the statistical significance of effects. The problem is especially severe for coefficients of higher-level variables, school characteristics in the present case. Second, a multilevel approach allows the investigation of the influence of school-level variance on pupil response, providing insight into whether school or pupil level factors are more important in explaining pupil response. Determining the relative importance of different levels gives important insights to the level 'at which the action lies' (Brown *et al.* (2005)). The variance partition coefficient (VPC) indicates the proportion of variation in pupil-level response behaviour that is due to the school. Different definitions of the VPC can be used for multilevel logistic models (see Goldstein *et al.* (2002); Goldstein (2011), Ch. 4.9). Here, the latent variable approach, sometimes referred to as the threshold model (Snijders and Bosker (2012)), is used.

Finally, the approach allows naturally for the exploration of contextual effects that we emphasised in the Introduction to be potentially important for pupil response. For example, we can test whether pupil response behaviour depends on the achievement levels of the school as a whole with higher achieving schools impacting positively on students' response probability. More fundamentally, we can investigate whether the impact of pupil level characteristics, such as individual ability, vary across schools (random coefficient models).

Let y_{ij}, the dependent variable, denote the outcome for pupil i in school j coded:

$$y_{ij} = \begin{cases} 1 & \text{response} \\ 0 & \text{non-response.} \end{cases}$$

Denoting the probability of response by $\pi_{ij} = \Pr(y_{ij} = 1)$, and taking non-response as the reference category, the two-level logistic model for student participation with a random intercept and random slope can be written as

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + \boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \lambda^T \boldsymbol{z}_j + u_{0j} + x_{1ij} u_{1j}, \tag{3}$$

where $\boldsymbol{x}_{ij}$ is a vector of student covariates and cross-level interactions between pupil and school level effects, x_{1ij} is a student level covariate with a random slope, $\boldsymbol{z}_j$ is a vector of school-level characteristics (which may differ from the one specified in (1)), $\boldsymbol{\beta}$ and λ are vectors of coefficients and u_{0j} and u_{1j} are random effects (random intercept and random slope

8

respectively), representing unobserved school effects. The random effects are assumed to follow a bivariate normal distribution, i.e. $\boldsymbol{u}_j = (u_{0j}, u_{1j})^T \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_u)$, where $\boldsymbol{\Omega}_u$ is a 2x2 covariance matrix of the form

$$\boldsymbol{\Omega}_u = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}$$

with variances σ_{u0}^2 and σ_{u1}^2 and covariance σ_{u01}. The intercept variance, σ_{u0}^2, represents the residual between-school variance in the log-odds of response versus non-response, while the slope variance, σ_{u1}^2, is the between-school variance in the effect of the covariate x_{1ij}. The covariance term, σ_{u01}, represents the covariance between the random intercepts and random slopes school effects and the two random effects, u_{0j} and u_{1j}, can be correlated (i.e. it is not assumed that $\sigma_{u01} = 0$).

The following modelling strategy is employed. First, random intercept models with focus on pupil-level non-response correlates are explored. Then contextual school level characteristics are added. Next cross-level interactions are explored. Finally random slopes on important pupil-level characteristics are investigated. The models are estimated using the xtlogit and xtmelogit command in STATA, version 12.0.

4. Results

4.1. Modelling school response

Given that the PISA target variable is achievement, we first investigate if school response is associated with pupil ability within schools. Figure 2 shows a small reversed u-shape relationship between the schools' average mean KS4 point score and school response using non-parametric local polynomial regression (Fan 1992). The graph is limited to values of achievement that are between the 10th and 90th percentile of the distribution. After an initial slight decrease of the response probability for schools with on average low achieving students, it increases to its maximum around the median average KS4 point score of 36 and decreases thereafter. Schools with low achieving and schools with high achieving students have both comparable low response rates and should therefore be equally represented in the responding school sample. This could suggest that school response does not contribute to bias of the PISA outcome variable. However, school response might well impact on the variance of the PISA achievement measure that is likely to be underestimated. The impact of non-response on the distribution of student achievement is estimated in Micklewright *et al.* (2012). It is however important to note, that the confidence intervals are large for all achievement values, with most of the intervals overlapping.

[Figure 2 about here]

9

In a second step, we model school response using logistic regression models. Table 3 shows the results for 5 nested models. Our first model is similar to the non-parametric regression with regard to variables included: we explain school response solely by including the schools' mean Key Stage 4 point score (KS4ps) and its square (divided by 100). Both coefficients are significant at the 5% level and indicate a turning point of response at the average achievement of 37.5 which fits relatively well to results displayed in Figure 2. For the 500 schools sampled for which Free School Meal (FSM) information is available, schools average pupil achievement and percentage of pupils with FSM eligibility is highly correlated (r=-0.68). The question therefore derives, how mean achievement and FSM eligibility are jointly related to schools' response probability. Model 2 in Table 3 indicates that conditional on percentage of pupils with free school meals in a school, pupils' mean achievement is not significant any more. This also does not change once we condition on whether the school is private or public (which is not significant) and the year of data collection (Model 3). The considerable size of the coefficient for FSM stays quite similar once we condition on whether a school was main, first or second replacement (Model 4). The estimated coefficient for FSM eligibility in Model 4 (-2.6) implies that for a school with a predicted probability of response of 0.5 a 10 per cent rise in free school meal eligibility decreases the probability of response by as much as 7 per cent (-2.6*0.25*0.1).

[Table 3 about here]

This result has a clear implication for the survey design. As discussed above, on the sampling frame schools are stratified by mean achievement. This ensures that replacement schools match initial non-responding schools in their average pupil ability. However, given that free school meal eligibility seems to be more important for response than achievement, free school meal eligibility would be an important factor to stratify on. This would ensure that first and second replacement schools would match the initial non-responding schools in terms of its free-school meal eligibility and therefore create a more representative school sample.

As discussed above, replacement schools have a much lower response probability. Model 4 shows that this is true even conditional on other explanatory variables. The estimated coefficients for first (-1.0) and second replacement schools (-1.5) in Model 4 imply that in comparison to a school with a predicted probability of response of 0.5, the response probability decreases by 25 per cent for a first and 38 per cent for a second replacement school. This high non-response of replacement schools would be of great concern, if replacement schools, for example, had a higher proportion of low performing students or more students eligible for FSM, which might be less likely to respond. In that case, the use of replacement schools would not, as intended, improve the sample representativeness but actually decrease it. We find that different

sets of interaction variables between replacement schools and mean achievement in schools are not significant. However, as results of model 5 indicate replacement schools with a higher proportion of children eligible for FSM are considerably less likely to respond (interaction effect significant at the 10% level). Since proportion of FSM students in the school is the most important explanatory factor of non-response (Table 3), this may question whether replacement schools help in improving the sample.

We also include region, proportion of males in school, single gender school and KS3 scores (sat at age 14) conditional on KS4 scores in our models (not shown). These variables are not found significant at the 10%. Similarly, any interaction of achievement with year cannot explain school response.

As often the case in social science applications and non-response modelling (Groves (2006)), the models explain only about 10 per cent of the variance in school-level response (pseudo R-square value) and we may therefore not be very successful in predicting non-response. This may indicate that other non-observed variables not in the data set, like characteristics and experience of headteachers and school governments, might impact on non-response, or that non-response has random elements.

[Table 3 about here]

4.2. Modelling pupil response

How is student response related to student ability? Figure 3 uses a non-parametric local polynomial regression in order to examine the association of KS4 point score with response probability. Students with low KS4 point score have a considerably lower response probability than children with higher ability. The increase in response probability is large and almost linear if we focus on the achievement distribution up to the 25th percentile, which is at the value of 28 KS4 points. The curve continues to increase at a lower rate up to the 75th percentile of the achievement distribution (48 KS4 points) and slightly decreases after that.

[Figure 3 about here]

Results for the logistic multilevel modelling are presented in Table 4. We first explore the impact of the main student achievement variable (KS4) on pupil response. Model 2 uses a quadratic form and Model 3 the natural logarithm of the centred KS4 score. We also experimented with spline regressions. We settle with the logarithmic form since it makes results of random slopes (Models 7 and 8) easier to interpret. However, results presented in Table 4 are robust to the choice of KS4 score specification.

[Table 4 about here]

In a first step, we examine the impact of student characteristics on pupil response. In line with non-parametric results we find a considerable impact of student achievement on response. This is conditional on the year the survey was conducted and gender; the latter coefficient indicating a significantly higher response probability for boys. Interestingly and in contrast to school response results, students' free-school meal eligibility cannot explain student response conditional on achievement (results not shown).

Across all multilevel models the random school variance is highly significant, indicating a clear school effect on pupil-level response. Using multilevel models we can measure the per cent of variation in student response due to differences between schools. Calculating the Variance Partition Coefficient (VPC) using the threshold model shows that this is a non-negligible number of 15% (Models 1 to 6). Figure 4 (based on Model 3) provides the estimates of school intercepts with the 95 per cent confidence interval indicating that random effects differ significantly for a considerable number of schools.

[Figure 4 about here]

Contextual effects at the school level are also investigated. Students are embedded in a school environment which might be less or more encouraging to sit the exam. We test this in Models 4 to 6 by including school-level variables but no significant effects are found. Pupils in private schools, in schools with a low proportion of FSM eligible students or in schools with high average achievement do not differ in their response behaviour to their counterparts in other schools. Also the geographic region of the school is not significant (result not shown). It is notable that our school variables are limited by providing merely information on pupil population within schools. Other school characteristics like schools' previous participation in surveys, school ethos, parents' background and attitudes and headteacher characteristics could well be associated with student non-response.

Even though school factors are on average not significant for explaining response, their association with student response might vary for different kinds of students. For example, since the PISA test took place during the preparation for GCSE exams, schools with low performing students (and hence at risk to be at a low position in the league table) might have encouraged especially higher ability students to participate in the survey while offering teaching sessions for those students refusing to take part. We therefore test for these cross-level interactions between pupil and school characteristics. Again using different specifications for achievement variables, we cannot find any significant results (Model 6). Pupil achievement impacts also similarly for schools differing by high and low social background of students (results not shown). The same is true for cross-level interaction of type of school and pupil achievement score (results not

shown). We therefore conclude that given the available variables neither school characteristics nor cross-level interaction of school with pupil characteristics are successful in explaining pupil non-response.

Up to now, we use random intercept models allowing the probability of response of students to depend on the school they attend. This is achieved by allowing the model intercept to vary randomly across schools. We assume however that the effect of, for example, individuals' achievement is the same across schools. In a final step, we relax this assumption and allow the random intercept and the slope on the achievement variable KS4 score to covary. Results for this random slope model are provided in Models 7 and 8.

Using the likelihood ratio test, Model 7, including a basic set of explanatory variables and the random slope, is considerably improved compared to Model 3 ($p<0.005$), containing the same set of basic variables but without the random slope. The effect of log achievement on the log-odds of response in school j is estimated as $0.827 + u_{1j}$. Simplified this means that the odds of response increase by about 8 per cent ($1.10^{0.827}$) if achievement increases by 10 per cent assuming the school intercept to be 0. The between-school variance in the effect of achievement is estimated as 0.135. Because the log of achievement has been centred at its sample mean, the intercept variance of 0.61 is interpreted as the between school variance in the log-odds of response at the mean of the log of the achievement score.

The most interesting result is the positive intercept-slope covariance estimate of 0.11 (significant at the 5 per cent level) (correlation coefficient is 0.384). This implies that schools with above average pupil response tend to also have an above average effect of achievement on response probability. Figure 5 shows this relationship by plotting the random slope for the centred achievement variable against the random intercept for schools. In schools with low proportion of students responding, the relationship between achievement score and response in the school is low. In schools with a high proportion of responding pupils achievement impacts more on response. In other words, schools with higher student response contribute more to the bias of the data then students with lower student response, since in the former achievement has a much higher explanatory power for non-response than in the latter schools.

[Figure 5 about here]

This effect may be difficult to explain, but shows clearly that higher student participation within the schools comes at the price of worse data quality of the PISA target variable achievement. It might be that there is a 'natural' level of students' interest to participate in the survey. Some schools accept this level and do not provide further encouragement for students to take part (schools with lower student response). Other schools might try to encourage more

students to sit the test thereby achieving only that better ability students can be changed in their opinion on participation.

Model 8 shows that adding school characteristics does not change findings on the random slope and again confirms the lack of school variables we can draw on in this study to explain the random school effect.

5.   **Conclusions and implications for survey practice**

Over the last 15 years educational research has been enriched by many international educational achievement surveys. The results of those have widely impacted on education policy debates as well as guided educational policy formulation in countries participating in the surveys. However, there is a considerable gap in the literature examining the quality of these data, in particular with respect to non-response rates and non-response biases. The nature of response at school and student level differs from that in household surveys, so that literature on non-response for household surveys is not sufficient to close the current gap of knowledge on non-response bias in achievement surveys. This paper examines non-response patterns at school and pupil level for the most prominent achievement survey PISA in England using logistic and multilevel logistic models.

The analysis benefits from unusually rich information about non-responding schools and pupils by exploiting data from two important administrative sources including students' test results of national examinations sat at a similar time to the PISA test. These national scores are highly correlated with students' PISA scores, which is the central piece of the PISA survey. Access to fully observed data that is highly predictive of a survey target variable is rare in the survey methods context.

Results of the school level response analysis show that first and second replacement schools have a considerably lower response probability than initially approached schools. There seems to be a trend that replacement schools are more likely to respond if pupils' mean achievement within the school is higher (only 10 per cent significance). This questions whether the inclusion of replacement schools improves the representativeness of 15 year olds in the sample in terms of their achievement.

Surprisingly, results indicate that school response does not depend so much on how pupils perform on average but rather on the socio-economic background of pupils within schools measured by free school meal eligibility. Assuming a 50% school response, every 10 per cent point increase in average free school meal eligibility within the school decreases its response

probability by 7 per cent. Interestingly, school-level characteristics, such as gender composition, region, type of school (private vs public) and school size cannot explain school non-response.

The results have important implications for the sampling design of PISA. Currently initial schools are matched with those replacement schools that are in closest neighbourhood on the sampling frame regarding their composition of type of school, size, region and mean achievement. However, our results show that what really matters for school response is the free school meal eligibility of students within schools. Hence, if replacement schools are used to improve the representativeness of the school sample, it seems more important to match initial and replacement schools on the socio-economic composition of their students than on any of the factors currently used.

School response patterns differ greatly from pupil response patterns indicating that any kind of survey design needs to consider different response mechanisms at both levels for achieving the best possible representative sample. In contrast to school response, students' ability is the strongest predictor of student response while students' free school meal eligibility does not gain any importance. Results of the multilevel models show that schools matter since 15 per cent of variation in students' response are due to differences between schools. Examining contextual effects with the aim of explaining such school-level influences, we do not find that students in private schools, schools with high achievement students and schools with low free school meal eligibility show different behaviour to their peers in other schools. However, it is possible that there are other school characteristics which could explain differences between schools on pupil response like characteristics of the head of the school, school ethos, parents' attitudes, the number of requests schools previously received to participate in surveys as well as their previous response behaviour.

Given the importance of achievement for pupil response, we tested whether the association between student ability and response differed by school environment using cross-level interactions and by schools using random slope models. While a variety of cross-level interactions proved not to be significant, we found that schools with higher pupil response tend to also have an above average effect of achievement on response probability. This clearly shows that higher student participation within the schools comes at the price of worse data quality of the PISA target variable on achievement. This means that existing practices to exclude schools with low student response from the dataset should be revisited.

Future research aiming to evaluate the quality of educational achievement surveys needs to investigate whether non-response patterns found for the UK, can be generalised over time and across countries.

15

References

Agresti, A.(2013) *Categorical Data Analysis*, 3rd edition. Hoboken, New Jersey: John Wiley and Sons.

Blatchford, P., Goldstein, H., Martin, C., and Browne, W. J., (2002) A study of class size effects in English school reception year classes, *Brit. Educ. Res. J.*, 28, 2, 169-185.

Brown, W.J., Subramanian, S.V., Jones, K. and Goldstein, H. (2005) Variance Partitioning in Multilevel Logistic Models that Exhibit Overdispersion. *J. R. Statist. Soc.* A, 168, 3, 599-613.

Dearden, L., Miranda, A. and Rabe-Hesketh, S. (2011), Measuring School Value Added with Administrative Data: The Problem of Missing Variables. *Fisc. Stud.*, 32: 263–278.

Durrant, G.B. and Steele, F. (2009) Multilevel Modelling of Refusal and Noncontact Nonresponse in Household Surveys: Evidence from Six UK Government Surveys, *J. R. Statist. Soc.* A, 172, 2, 361-381.

Durrant, G.B., Groves, G., Staetsky L. and Steele, F. (2010) Effects of Interviewer Attitudes and Behaviours on Refusal in Household Surveys. *Public Opin. Quart.*, 74, 1, 1-36.

Fan, J. (1992) Design-adaptive nonparametric regression. *J. Am. Stat. Assoc.*, 87, 998-1004.

Gelman, A. and Hill, J. (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge, Cambridge University Press.

Goldstein, H. (2011) *Multilevel Statistical Models.* 4th ed., Chichester: Wiley & Sons.

Goldstein, H., Browne, W. J. and Rasbash, J. (2002) Partitioning Variation in Multilevel Models. *Understanding Statistics*, 1, 223–231.

Gray, J., Goldstein, H. and Thomas S. (2001) Predicting the future: the role of past performance in determining trends in institutional effectiveness at A level, *Brit. Educ. Res. J.*, 19:24, 3417-3432.

Groves, R. M. (2006) Nonresponse rates and nonresponse bias in household surveys. *Public Opin. Quart.*, 70, 646–75.

Hawkes, D. and Plewis, I. (2006) Modelling non-response in the National Child Development Study, *J. R. Statist. Soc.* A, 169, 479–491.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T. M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R. M. and Raghunathan, T. E. (2010) Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. *J. R. Statist. Soc.* A, 173, 389–407.

Leckie, G. and Goldstein, H. (2011) Understanding Uncertainty in School League Tables. *Fisc. Stud.*, 32: 207–224.

Little, R. and Vartivarian, S. (2005) Does weighting for nonresponse increase the variance of survey means?, *Surv. Methodol.*, 31, 161–168.

Micklewright, J. , Schnepf, S. V. and Skinner, C. J. (2012) Non-response biases in surveys of school children: the case of the English PISA samples. *J. R. Statist. Soc.* A, 175 (4), 915-938.

Organisation for Economic Co-operation and Development (2004) *Learning for Tomorrow'sWorld: First Results from PISA 2003.* Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development (2012). *PISA 2009 Technical Report.* Paris: OECD Publishing

Olson, J. F., Martin, M. O., Mullis, I. V. S., Mullis, I. V. and Arora A. (eds.) (2008) *TIMSS 2007 Technical Report.* Boston and MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and IEA TIMSS & PIRLS.

Pickery, J., Loosveldt, G. and Carton, A. (2001) The effects of interviewer and respondent characteristics on response behavior in panel surveys. *Sociol. Meth. Res.*, 29, 509–523.

Pike, G. R. (2008) Using Weighting Adjustments to Compensate for Survey Nonresponse. *Res. High. Educ.*, 49(2), 153–171.

Porter, S. R., and Whitcomb, M. E. (2005) Non-response in student surveys: the role of demographics, engagement and personality. *Res. High. Educ.*, 46(2), 127–152.

Rasbash, J., Leckie, G.B., Jenkins, J. and Pillinger, R.J. (2010) Children's educational progress: partitioning family, school and area effects, J. R. Statist. Soc. A, 173(3), 657-682.

Rust, K., Krawchuk, S. and Monseur, C. (2013) PISA Student Nonresponse Adjustment Procedures. In *Research on PISA New York* (ed M. Prenzel), pp. 87–102. Berlin: Springer.

Snijders, T. and Bosker, R. (2012) *Multilevel Analysis, An Introduction to Basic and Advanced Multilevel Modelling.* London: Sage Publications.

Steinhauer, H.-W. (2014) Modelling unit nonresponse in educational multi-informant studies: a multilevel approach using bivariate probit models. *Poster presented at the Panel Survey Methods Workshop*, Ann Arbor.

Sturgis, P., Smith, P. and Hughes, G. (2006) A study of suitable methods for raising response rates in school surveys. *Research Report 721*. Department for Education and Skills, London.

Tables and Figures

Table 1. Approached schools, by replacement status and response outcome

	2000		2003		Total	
	Total # schools	% responding	Total # schools	% responding	Total # schools	% responding
Initial	179	60.3	139	73.4	318	66.0
First	78	42.3	40	37.5	118	40.7
Second	45	24.4	28	35.7	73	28.8
Total	302	50.3	207	61.4	509	54.8

Note: the total number of responding schools is 279.

Table 2. Sampled pupils, by response outcome

	2000		2003		Total	
	Total # pupils	% responding	Total # pupils	% responding	Total # pupils	% responding
Total	4,846	81.0	4,032	71.9	8,878	76.8.

Note: the total number of responding pupils is 6,821.

Table 3. Results of 5 nested logistic models for school response

	(1)	(2)	(3)	(4)	(5)
KS4 ps	0.165**	0.0221	0.0467	-0.0115	-0.0235
	(0.0802)	(0.0940)	(0.101)	(0.102)	(0.103)
KS4 ps^2 / 100	-0.220**	-0.0770	-0.102	-0.0223	-0.012
	(0.103)	(0.114)	(0.127)	(0.126)	(0.126)
FSM proportion		-2.838***	-2.467**	-2.597**	-2.017*
		(1.004)	(1.015)	(1.016)	(1.058)
Missing FSM info		-0.802	-1.048	-0.737	-0.771
		(0.748)	(0.824)	(0.800)	(0.785)
Private school			-0.00401		
			(0.479)		
First replacement				-1.027***	-0.675**
				(0.227)	(0.306)
Second replacement				-1.541***	-1.163***
				(0.290)	(0.364)
Size in 100				0.208	0.212
				(0.136)	(0.137)
Year 2000 school			-0.438**	-0.280	-0.279
			(0.194)	(0.209)	(0.209)
FSM replacement school					-2.708*
					(1.620)
Constant	-2.755*	0.882	0.545	1.556	1.7410
	(1.524)	(1.974)	(2.061)	(2.058)	(2.070)
Pseudo R-square	0.00701	0.0206	0.0282	0.0936	0.0979
log-likelihood	-348.0	-343.2	-340.6	-317.6	-316.1

Note: school sample size is 509, standard errors are reported in parentheses,
FSM = Free School Meal, KS4 ps = Key Stage 4 point score
*** denotes p<0.01, ** p<0.05, * p<0.1.

**Table 4. Estimated coefficients of multilevel logistic models for student response**

		Random intercept models						Random intercept and slope models	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Student characteristics	Ln(Ks4 ps), centred			0.776***		0.782***	0.763***	0.827***	0.816***
				(0.042)		(0.042)	(0.048)	(0.052)	(0.058)
	Ks4 ps, centred		0.022***						
			(0.002)						
	Ks4 ps²/ 100, centred		-0.117***						
			(0.011)						
	Male		0.184**	0.188**		0.192**	0.191**	0.190**	0.192**
			(0.059)	(0.058)		(0.059)	(0.059)	(0.059)	(0.060)
School characteristics	Private school				0.400	0.441			
					(0.313)	(0.314)			
	% FSM in school				0.270	-0.118	0.147		0.0206
					(0.645)	(0.664)	(0.659)		(0.661)
	Missing FSM				-0.930	-0.586	-0.296		-0.316
					(0.582)	(0.593)	(0.563)		(0.615)
	% male in school				0.135	-0.220	-0.0763		-0.0570
					(0.328)	(0.336)	(0.323)		(0.337)
	Mean Ks4 ps, centred				0.0236	-0.0134	-0.00204		-0.00408
					(0.013)	(0.013)	(0.012)		(0.012)
	Year 2000		0.499***	0.490***		0.485***	0.486***	0.493***	0.494***
			(0.109)	(0.109)		(0.112)	(0.113)	(0.110)	(0.114)
	Mean student x mean school Ks4 ps /100, centred						-0.604		-0.758
							(0.712)		(0.827)
	Constant	1.347***	-1.290***	1.040***	1.230***	1.143***	1.071***	1.055***	1.090***
		(0.056)	(0.088)	(0.084)	(0.203)	(0.205)	(0.200)	(0.0853)	(0.206)
Random components	VPC	0.159***	0.153***	0.152***	0.151***	0.150***	0.151***		
		(0.016)	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)		
	σ_{u0}	0.790***	0.770***	0.768***	0.764***	0.763***	0.766***		
		(0.047)	(0.047)	(0.047)	(0.046)	(0.047)	(0.047)		
	σ_{u0}^2							0.613***	0.610***
								(0.076)	(0.075)
	σ_{u1}^2							0.135**	0.137**
								(0.051)	(0.052)
	σ_{u01}							0.109**	0.105**
								(0.047)	(0.047)
	Log likelihood	-4557.2	-4337.1	-4350.1	-4550.3	-4348.7	-4349.3	-4341.1	-4340.2

Note: results are based on 8,878 students in 279 schools. Standard errors in parentheses;
KS4 ps = Key Stage 4 point score, FSM = free school meal
***denotes p<0.01, **p<0.01, *p<0.05.

Figure 1. Scatterplot of the PISA reading score against Key Stage 4 total points score: responding pupils, 2000 and 2003
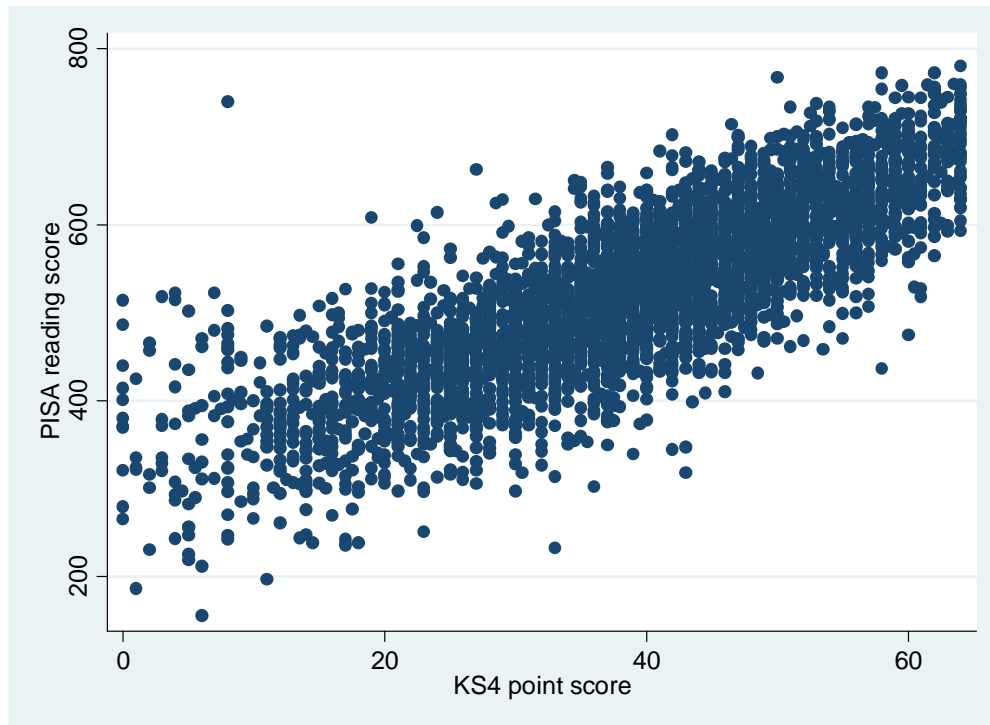
Note: correlation coefficient is 0.81; based on 3,923 students who sat the PISA reading test.

Figure 2. Response probability of school by pupils' mean total point score in Key Stage 4 achievement within schools, for pooled 2000 and 2003 data, using non-parametric local polynomial regression.
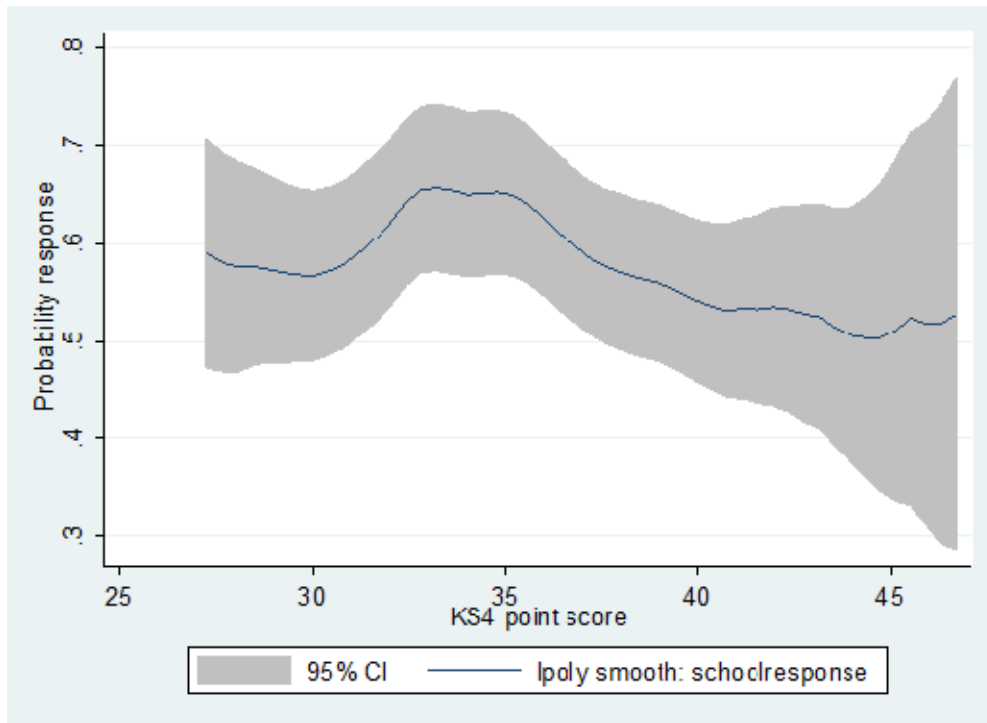
Note: average achievement score values between 10th and 90th percentile are taken into account.

Figure 3. Response probability of pupils by pupils' total Key Stage point score, for pooled 2000 and 2003 data using non-parametric local polynomial regression
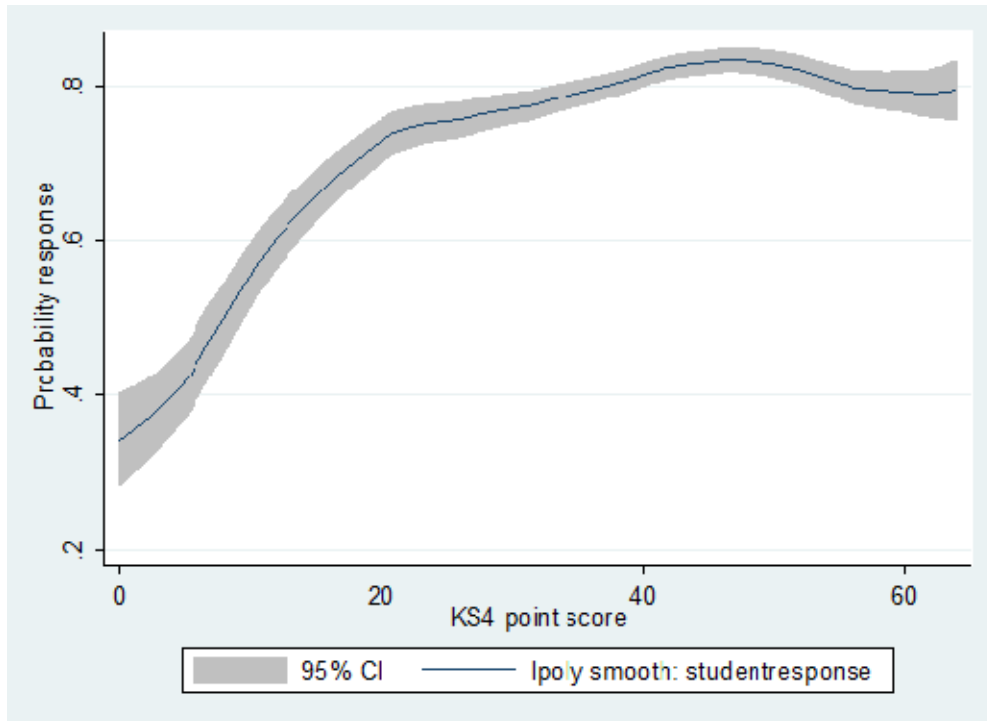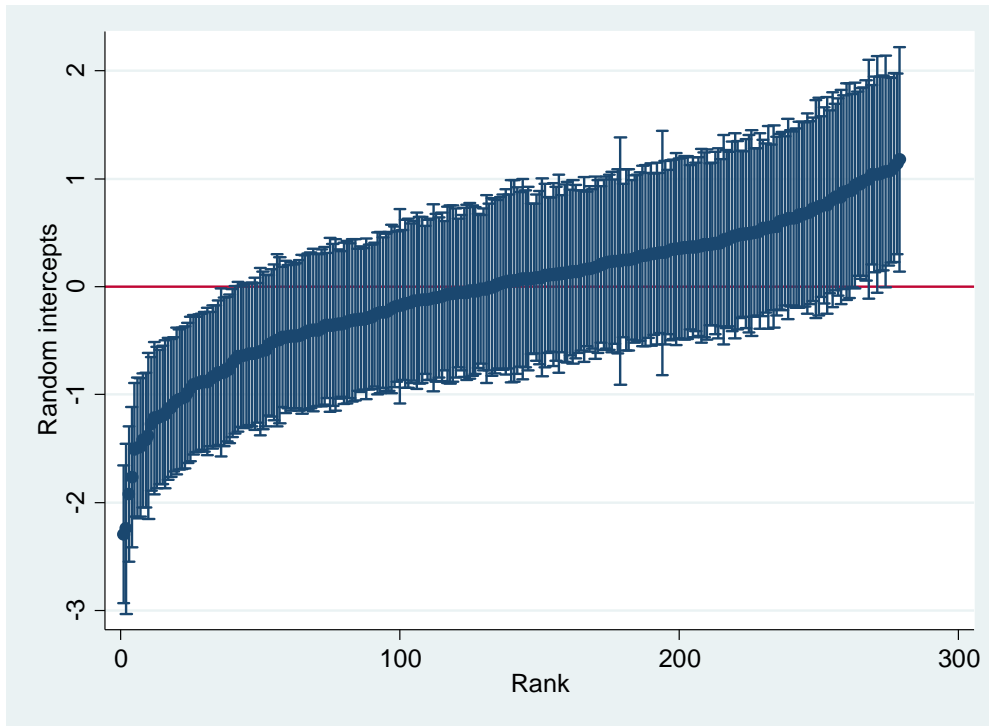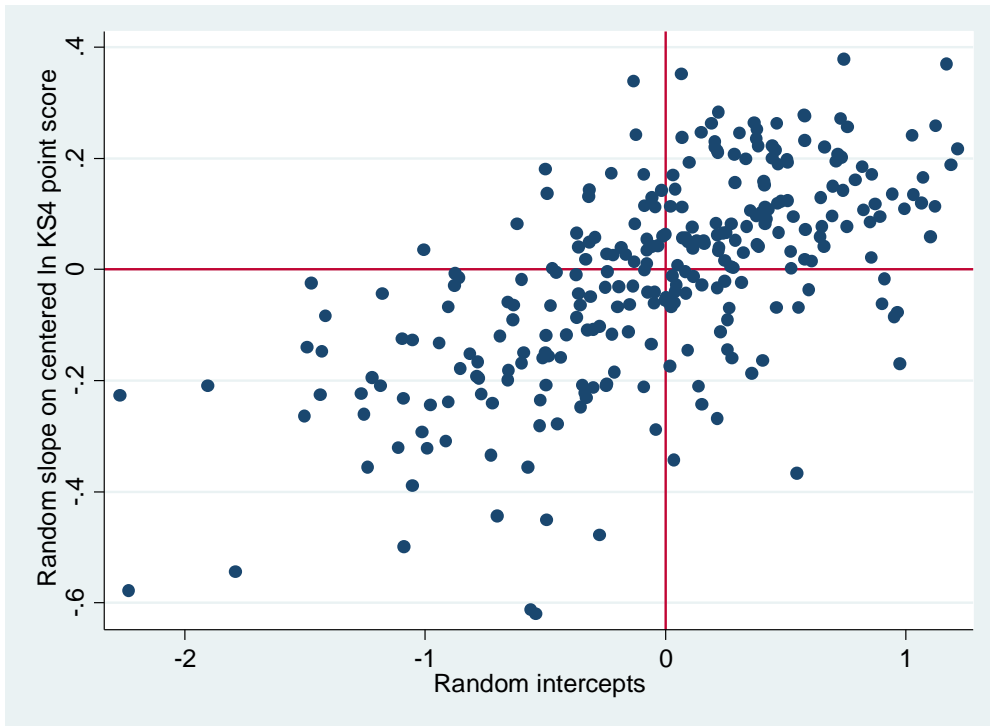
Figure 4. Estimates of school random intercepts from multilevel logistic model for pupil response (for model 3 in Table 4)

Note: the graph shows the estimates of the schools' random intercepts obtained in model 3 of Table 4.

Figure 5. Scatterplot of random intercept and random slope residuals for schools (based on Model 7 in Table 4)

Note: results are based on Table 4, Model 7.