An updated version of this manuscript has been accepted by the Journal of the Royal Statistical Society, Series A, Sept 2017

Which schools and pupils respond to educational achievement surveys? A focus on the English PISA sample

Gabriele B. Durrant[*] and Sylke V. Schnepf[+]
[*]Department of Social Statistics and Demography, University of Southampton, UK
[+]European Commission, DG Joint Research Centre, Italy

Summary
Non-response is a major problem facing research in the social sciences including in education surveys. Hence, research is needed to better understand non-response patterns as well as non-response as a social phenomenon. Findings may contribute to improvements in the future designs of such surveys. Using logistic and multilevel logistic modelling we examine correlates of non-response at the school and pupil level in 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, from two large-scale administrative sources on pupils' socio-economic background and results in national public exams. This information correlates highly with the PISA target variable. Findings show that characteristics associated with non-response differ between the school and pupil levels. Our results also indicate that schools matter in explaining pupil level response, which is often ignored in non-response analysis. Our findings have important implications for future education surveys. For example, if replacement schools are used to improve response, our results suggest that it may be more important to match initial and replacement schools on the socio-economic composition of their pupils than on any of the factors currently used.

Keywords: non-response patterns and processes; survey design; educational achievement survey; Programme for International Student Assessment (PISA).

Address for correspondence: Gabriele B. Durrant, Department of Social Statistics and Demography, School of Social Sciences, University of Southampton, email: g.durrant@southampton.ac.uk.

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 debates 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 (Araujo *et al.*, 2017), in particular with respect to non-response. Also, response in many education surveys has been falling (Sturgis *et al.* 2006) and questions arise about the underlying mechanisms of the non-response processes. In particular, it is of interest to better understand non-response, both non-response patterns and correlates of non-response. A better understanding will benefit future survey designs, in particular when variables about the survey data collection process (so-called paradata, Couper, 1998) are available. The sparse analysis of non-response in education surveys stands in contrast to household surveys, for which considerable research on the nature and correlates of non-response, including paradata, is available (for a summary see Groves, 2006; and also Durrant and Steele, 2009; Kreuter, 2013; Durrant *et al.,* 2015). This paper aims to address this shortcoming.

The most prominent achievement surveys are the Programme for International Student Assessment (PISA) (Organisation for Economic Co-operation and Development (OECD) (2012)), the Trends in International Maths and Science Study (TIMSS), and the Progress in International 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 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, their parents' permission to take part, the level of their schools' commitment to the survey and peer influences. For these reasons, response to this type of survey deserves more attention from researchers than it has received to date. Previous studies that analyse non-response in education data are limited but include Steinhauer (2014), Rust *et al.* (2013), Pike (2008), Porter and Whitcomb (2005) and Micklewright *et al.* (2012). Besides Micklewright *et al.* (2012) we are not aware of any literature analysing patterns of non-response in cross-national education surveys.

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 in England using two external administrative data sources: the Pupil Level Annual School Census and the National Pupil Database. In particular, 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 that provides information on non-respondents and allows non-response analysis (see also for example Dearden *et al.*, 2011; Micklewright *et al.*, 2012). Our access to administrative registers that contain information on results in national public exams is unique. Exam results are strongly correlated with the main PISA target variables that measure learning achievement – a characteristic essential 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 – another key advantage of the data.

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, with the aim of better understanding the non-response process and correlates of non-response. 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 low response rates (OECD, 2004). PISA's main aim is to assess pupils' learning achievement. Hence, a particular focus of this paper is the relationship of school and pupil response and pupil ability – mean pupil ability in the school in the case of school response and the individual's own ability for pupil response.

Although the focus is on correlates of non-response, these relationships in fact provide an indication of the extent of the risks of non-response bias in the estimates of learning achievement that the survey provides. If non-response is associated with ability as measured by the administrative data (even after controlling for other variables) then estimates for the survey outcomes on ability are likely to be biased. A companion paper estimates the likely sizes of these biases using different weighting methods, a subject we do not consider here (Micklewright *et al.*, 2012). Micklewright *et al.* (2012) used a part of the same dataset but focused on a smaller set of variables. In addition, they employed a single-level logistic model for pupil response, which was sufficient for the purpose of constructing non-response weights. They did not consider school response and analysed the 2000 and 2003 data separately.

In the present paper, we provide much more detailed modelling of both school and pupil non-response with the aim of better understanding the non-response process. Our models explore a richer set of explanatory variables, including available paradata. At the pupil stage we allow for the influence of the school on pupils' response behaviour, including random school effects and contextual school-level explanatory variables. We quantify this influence using a multilevel model. We are particularly interested in the influence of ability on the two types of non-response, as well as socio-economic indicators. Employing cross-level interactions between pupil and school characteristics we can, for example, examine whether high performing pupils in a low performing school are less likely to respond than high performing pupils in a high performing school. Furthermore, we use random slope models to relax the commonly made assumption that correlates of non-response do not depend on higher levels. Here we assess the school influence on the most important pupil response predictor: pupil ability.

The dataset also includes a small amount of paradata that describe the survey data collection process, such as replacement school indicator. Replacement schools, common in education data, are used as a means of increasing response rates with the hope of also subsequently reducing non-response bias. However, Sturgis *et al.* (2006) question whether replacement schools have the ability to reduce the risk of non-response bias or rather increase it, given the differences of fieldwork characteristics between the initial and replacement samples. This paper sheds further light on this question.

The 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 implement 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 (see also OECD, 2002, 2005 and 2014). 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. A more detailed description is given in Micklewright *et al.* (2012) and we hence limit the description here to only those sample design features relevant for this analysis. 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 pupils 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' (local education authority), '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 international surveys on children's learning achievement, the PISA sample design uses a system of 'replacement' of non-responding schools. (For an extended discussion on the use of replacement schools see Prais (2003), Adams (2003) and 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. The number of approached schools in England, including the replacement schools, was 302 in 2000 and 207 in 2003 (excluding any duplicate schools). Very similar designs for collecting the sample are used in all other countries of PISA. All countries (apart from the Russian Federation) use a two-stage stratified sample design. However, the choice of stratification variables is down to the countries participating in PISA and hence can differ from England. For example, in England private school is used as a stratification variable.

[Table 1 about here]

Table 1 shows that response rates for initial schools in England are 60% in 2000 and 73% in 2003. However, the majority of replacement schools do not respond, i.e. for the two years taken together only 41% of first and 29% of second replacement schools respond. These schools face a higher survey burden. They firstly have less time to conduct the survey and secondly the timing of the PISA survey data collection in 2000 and 2003 was closer to a nationally organised public exam, the results of which are used for school league tables in England.

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 pupils responded (3 schools) which we include. (Nowadays schools with less than 50% pupil response are excluded.) For the analysis here, we merge 2000 and 2003 data in order to increase the sample size. We hence 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). (Extensive analysis was also carried out on both datasets independently and all conclusions remained the same. Descriptive analyses of key variables are provided in the online Appendix, Table A1, and further information is available in the PISA technical reports (OECD, 2002 and 2005)).

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 pupils are those who did not take the PISA test, for example due to parents' lack of consent, other reasons like illness or the pupil not wanting to take part. Table 2 shows that 77% of pupils 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 resulting survey estimates will be biased. It follows that any kind of response modelling in the survey ideally needs to incorporate good auxiliary information on children's ability. England is unusual in having nationally organised exams for 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 considered 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 settled on the Key Stage 4 point score (KS4 ps) for the best 8 subjects, a variable often used in education research in the UK.

A scatterplot of the PISA reading score against the KS4 total points score shows that our auxiliary achievement measure is highly correlated with the PISA test score for reading among responding pupils (r=0.77) (see Figure A1 in the online Appendix). The correlation coefficients of our auxiliary variable with PISA math test score and with PISA science score are both 0.75 (based on 5,055 responding pupils for maths and 5,051 responding pupils for reading) – results not shown. (Note that not all pupils take all tests which is why the total numbers here are slightly less than the total numbers of responding pupils.) Our analysis therefore greatly benefits from access to fully observed auxiliary information that is highly associated with PISA achievement measures.

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 proportion of individuals receiving the benefit in the child's school.

We have access to the auxiliary information just described for the population of 15 year olds eligible for PISA. 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 over the PISA target population within schools: mean KS4 score, proportion of pupils with free school meals and proportion of male pupils 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 linkage used unique school and pupil identifiers that ONS and the Trust had added to the files. Linkage rates are high with a percentage loss due to non-matches of only around 1% at the school level for both years (2000 and 2003). The percentage loss for pupil level linkage is a little higher at 6% and just under 4%

for 2000 and 2003 respectively. Further details of the files and linkages are given in Micklewright *et al.* (2012). Besides FSM eligibility, missing rates on all other auxiliary variables at the pupil level are below 1%. At the school level the proportion of pupils eligible for FSM is missing for just under 2% of schools, which is taken into account in our modelling. (An indicator variable indicating if FSM is missing is included in the models and not significant throughout. Where the FSM proportion was missing it was set to the mean proportion.)

3. Methodology

We employ a sequential modelling approach, where we first model response at the school level using a (single-level) logistic model, analysing the effects of school characteristics on the school response propensities. Then for responding schools, we model pupil level response within schools 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 is modelled (Pickery *et al.*, 2001; Durrant and Steele, 2009; Durrant *et al.*, 2011).

To aid interpretation of coefficients in the logistic model for school response, we estimate the marginal effect of a one-unit change in a school-level characteristic z_k on the probability of response p as

$$\frac{dp}{dz_k} = p(1 - p)a_k , \tag{1}$$

where a_k is the coefficient of explanatory variable z_k and p is set by the researcher, most commonly at the average response probability estimated by the model (Gelman and Hill, 2006). The maximum marginal effect occurs where p is equal to 0.5.

To investigate non-response at the pupil level a multilevel modelling approach is employed, recognising the clustering of pupils within schools. Multilevel modelling is nowadays a well-established modelling approach and has become very popular in the education literature to analyse pupil and school effects on educational attainment. 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).

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) (Goldstein, 2011). Second, a multilevel approach allows the investigation of school effects on pupil response, providing insight into whether school or pupil level factors are more important in explaining pupil response. Determining the relative importance of factors at different levels gives important insights to the level 'at which the action lies' (Brown *et al.*, 2005). The variance partition coefficient (VPC) is interpreted as the proportion of variation in the underlying pupil-level response propensity that is due to differences between schools. (Different definitions of the VPC can be used for binary response models, see Goldstein *et al.*, 2002 and Goldstein, 2011, Ch. 4.9). Here, the underlying latent variable approach, sometimes referred to as the threshold model (Snijders and Bosker, 2012, p. 305), is used. Finally, the multilevel approach naturally allows for the exploration of contextual effects that we emphasised in section 1 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 pupils' response probability. More fundamentally, we can investigate whether the impacts of pupil level characteristics, such as individual ability, vary across schools (using random coefficient models).

Let y_{ij} denote the response 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 $p_{ij} = \Pr(y_{ij} = 1)$ a general two-level random coefficients logistic model for pupil participation can be written as

$$\text{logit}\,(p_{ij}) = \boldsymbol{b}^T \boldsymbol{x}_{ij} + \boldsymbol{u_j}^T \boldsymbol{w_{ij}}, \qquad (2)$$

where $\boldsymbol{x}_{ij}$ is a vector of pupil and school level covariates and their interactions and $\boldsymbol{w_{ij}}$ is a subset of pupil-level components of $\boldsymbol{x}_{ij}$ with random coefficients $\boldsymbol{u_j}$ at the school level. (More specifically, for a given element of $\boldsymbol{w_{ij}}$ we use the notation u_{0j}, u_{1j} and s_{u01} for the random intercept, random slope and covariance between random intercept and random slope effects respectively).

To test the significance of the random effects parameters and to allow comparison between nested models, the likelihood ratio test statistic is used (for variance components, since the null hypothesis is on the boundary of the parameter space, this is based on a 50:50 mixture of c^2 distributions; see Rabe-Hesketh and Skrondal, 2012a and 2012b, p. 88-89; pp. 88-89; p. 536; Verbeke and Molenberghs, 2000, ch. 6.3; Zhang and Lin, 2008, pp. 22-25).

We employ the following systematic modelling strategy. First, we explore random intercept models with a focus on pupil-level non-response correlates. Then, we add contextual school level characteristics to explore the influence of schools on pupil-level response. Next, we test for cross-level interactions. Finally, we investigate random slopes for important pupil-level characteristics. Random slope and random intercept effects are allowed to co-vary. We also carried out comprehensive sensitivity checks where the order of variables entered was varied, but the conclusions remained the same.

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 1 shows a weak 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 pupils, it increases to its maximum around the median KS4 point score of 36 and decreases thereafter. Schools with on average low achieving pupils and schools with on average high achieving pupils have comparable low response rates and schools with middle range achievement scores have the highest response rate. This relationship between achievement and response motivates the inclusion of a quadratic term of achievement in later models. It also indicates that the variance of the PISA achievement score is likely to be underestimated in the presence of such non-response patterns. (Schools with pupils at the lower and at the higher end of achievement scores are more likely to not respond, reducing the variance of the observed achievement scores). This effect is important to note since, as a result, educational disparity in England is likely to be underestimated. It is however important to note that the confidence intervals are large for all achievement values.

[Figure 1 about here]

We model school response using logistic regression models. Table 3 shows the results for 5 nested models, exploring the inclusion of a range of variables and interaction effects. 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 (KS4 ps) 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, consistent with the pattern shown in Figure 1. For the 500 schools sampled, for which free school meal (FSM) information is available, average pupil achievement and proportion of pupils with FSM eligibility are highly correlated (r=-0.68). How are mean achievement and FSM eligibility jointly related to schools' response probability? Model 2 in Table 3 indicates that, conditional on proportion of pupils with free school meals in a school, pupils' mean achievement is no longer significant. This 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). (In Models 1-3 school size was also included but not found significant.) 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 percentage point rise in free school meal eligibility decreases the probability of response by as much as 6.5 percentage points (-2.6*0.25*0.1) (following equation 1). Indicators of whether a school was first or second replacement (Model 4) are significant (at 5% and 1% levels) impacting negatively on response behaviour. Model 4 shows that the replacement school effects remain significant after conditioning 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 percentage points for a first and 38 percentage points for a second replacement school.

The lower response propensity of replacement schools might be due to the shorter time span these schools have for conducting the survey. This could also explain why second replacement schools are less likely to respond than first replacement schools. Furthermore, the high non-response of replacement schools would be of concern if replacement schools, for example, had more pupils eligible for FSM – a variable not used for stratification but associated with response. 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 between FSM and replacement school is significant at the 10% level). To interpret this effect further, we also consider the effect of a 10 percentage point increase in FSM from the mean on the marginal probability that a school responds, setting all other variables to their means. We find that this decreases school response by 7.5 (=0.548-0.473) percentage points. Since the proportion of FSM pupils in the school is an important explanatory factor of non-response (Table 3), this may call into question whether replacement schools actually help in improving the data quality of the sample.

[Table 3 about here]

The results have clear implications for the survey design. Stratification aims at improving the precision of the sample by choosing stratification variables that are highly correlated with the outcome variable educational achievement. As discussed above, schools are stratified by mean achievement on the sampling frame. Besides decreasing sampling variation, 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 and is also correlated with achievement, free school meal eligibility would be an important stratification variable to add. This would ensure that first and second replacement schools would match the initial non-responding schools in terms of their free-school meal eligibility and therefore create a more representative school sample. Further, using replacement schools for balancing response may not be very useful and in fact counterproductive, leading to a lower representation of schools with higher proportion of FSM pupils, with the potential to increase effects of non-response on survey outcomes. We also include region, proportion of males in school, an indicator of single gender school and mean KS3 scores (on tests taken at age 14) conditional on mean KS4 scores in our models (not shown). These variables are not found to be significant at the 10% level. Similarly, the interaction of achievement with year does not explain school response. Overall, all 5 nested models show successive reductions in the log-likelihood value (see Table 3), indicating that Model 5 is the best fit with not much difference between models 4 and 5 but with considerable reductions in comparison to models 1 to 3.

4.2. Modelling pupil response

We now turn to investigating pupil level response. We first explore to what extent pupil response is related to pupil ability. Figure 2 uses a non-parametric local polynomial regression in order to examine the association of KS4 point score with response probability. Pupils with a 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 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 decreases slightly after that.

[Figure 2 about here]

Results for the multilevel logistic modelling are presented in Table 4. We first explore the impact of the main pupil 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 considered 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 pupil characteristics on pupil response. In line with the non-parametric results we find a considerable impact of pupil achievement on response. This is conditional on the year the survey was conducted and gender; the latter coefficient indicates a significantly higher response probability for boys. Interestingly, and in contrast to school response results, pupils' free-school meal eligibility does not explain pupil response conditional on achievement (results not shown).

Across all multilevel models the between school variance is highly significant, indicating a clear school effect on pupil-level response (for example for Model 1, the empty random intercepts model, the likelihood ratio test statistic for a test of the between-school variance is 497.2 with a p-value <0.001, which strongly suggests that a multilevel model is required). Using multilevel models we can measure the per cent of variation in pupils' underlying response propensities due to differences between schools. Calculating the Variance Partition Coefficient (VPC) using the threshold model shows that this is a non-negligible 15% (Models 1 to 6), and this effect remains stable across all models.

Contextual effects at the school level are then investigated. Pupils are embedded in a school environment which might be more or less encouraging to participate in PISA. 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 pupils or in schools with high average achievement do not differ in their response behaviour relative to their counterparts in other schools. Nor is the geographic region of the school significant (result not shown). We also explored the effect of the within-school variance of achievement score. We explored both the variance based on the original variable KS4 and the log of achievement, but the effects are not significant in any of the models considered.) It is notable that our school variables are limited by providing merely information on pupil populations 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 pupil non-response.

Even though the measured school factors are on average not significant for explaining response, their association with pupil response might vary for different kinds of pupils. For example, since the PISA test took place during the preparation for GCSE exams, schools with low performing pupils (and hence at risk to be in a low position in a league table) might have encouraged especially higher ability pupils to participate in the survey while providing normal teaching for those pupils refusing to take part. We therefore test for these cross-level interactions between pupil and school characteristics but, again using different specifications for achievement variables, we do not find any significant effects (Model 6). Pupil achievement impacts on response similarly for schools differing by high and low social background of pupils (results not shown). The same is true for the cross-level interaction of type of school and pupil achievement score (results not shown). We therefore conclude that given the available variables, neither observed school characteristics nor cross-level interactions of school with pupil characteristics are successful in explaining pupil non-response.

Up to now, we used random intercept models allowing the probability of response of pupils to depend on the school they attend. This is achieved by allowing the model intercept to vary randomly across schools. This assumes, however, that the effect of, for example, individuals' achievement is the same across schools. To assess the school influence on the most important pupil response predictor, pupil ability, we test for a random slope on achievement. (We 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.

Model 7 - including a set of explanatory variables and the random slope - is a significant improvement on Model 3 containing the same set of variables but without the random slope (the likelihood ratio test statistic is 18.0, $p<0.001$). The model with the random slope for achievement implies that the between-school variance depends on achievement and the variance increases with achievement. The between-school variance in the log-odds of response is estimated as 0.613+2*0.109 achievement+0.135 achievement^2 (see also Rabe-Hesketh and Skrondal, 2008, p. 150). The effect of log achievement on the log-odds of response in school j is estimated as $0.827 + u_{1j}$. This implies that the odds of response increase by about 8 per cent ($1.10^{0.827}$) for a 10 percentage point increase in achievement setting the school slope effect u_{1j} to 0. The between-school variance in the effect of achievement is estimated as 0.135. Hence, a 95% coverage interval for the school slope on achievement is estimated as 0.827±1.96*sqrt(0.135)= 0.107 to 1.547. Thus, we would expect the middle 95% of schools to have a slope between 0.107

to 1.547 (see also Rabe-Hesketh and Skrondal, 2008, p. 153). 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; the 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 A3 in the online appendix shows this relationship by plotting estimates of the random slopes for the centred achievement variable against the random intercepts for schools.) In schools with a low proportion of pupils responding, the relationship between achievement score and response in the school is weak. In schools with a high proportion of responding pupils achievement impacts more on pupil response. In other words, schools with higher pupil response contribute more to the non-representativeness than pupils with lower pupil response, since in the former achievement has a higher explanatory power for non-response than in the latter schools.

This effect may be difficult to explain, but shows clearly that higher pupil 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 pupils' interest to participate in the survey. Some schools accept this level and do not provide further encouragement for pupils to take part (schools with lower pupil response). Other schools might try to encourage more pupils to participate in PISA, thereby resulting in only better ability pupils changing their attitude to participation.

Model 8 shows that adding school characteristics does not change findings related to the random slope and again confirms the lack of school variables we can draw on in this study to explain the differences among schools in the pupil response probability.

5. Conclusions and implications for survey practice of PISA

Over the last 15 years educational research has been enriched by many international educational achievement surveys. The results of these surveys have impacted importantly on education policy debates and have 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. The nature of response at school and pupil levels differs from that in household surveys, which means that the literature on non-response for household surveys is not sufficient to close the current gap of knowledge on non-response in achievement surveys. This paper examines non-response patterns and correlates at school and pupil levels for the most prominent achievement survey (PISA) in England to better understand the non-response process and non-response as a social phenomenon. A better understanding will guide improvements to survey designs and survey practice.

Non-response analysis is often restricted by the absence of fully observed data. Our analysis benefits from unusually rich information about non-responding schools and pupils by exploiting data from two important administrative sources, including pupils' test results of national examinations taken at a similar time to the PISA test. These national scores are highly correlated with pupils' PISA scores. Access to fully observed data that is highly predictive of a survey target variable is rare.

School-level response and implications for survey practice

Results of the school level response analysis show that first and second replacement schools have a considerably lower response probability than initially approached schools even after controlling for other factors, such as ability and type of school. Slightly 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 as measured by the proportion eligible for free school meals. In particular, the analysis found a significant interaction between FSM eligibility and replacement schools, indicating that replacement schools with a higher proportion of pupils eligible for FSM are significantly less likely to respond than replacement schools with a lower proportion of pupils eligible for FSM. A reason for this might be that schools in socially deprived areas in the light of associated social problems and education standards may set other priorities than responding to a cross-national education survey. Moreover, national public exams in 2000 and 2003 were taken very close to the PISA tests which may have also increased the chances of non-response. Interestingly, school-level characteristics, such as gender composition, region, type of school (private vs public) and school size do not explain school non-response.

These results have important implications for the sample design of PISA. Currently initial schools are replaced with the school that is most similar in terms of type of school, size, region and mean achievement. While stratification by these variables is sensible in terms of increasing precision of the estimate since these variables are highly correlated with the outcome variable achievement, our results show that such variables are not important for non-response. Instead, our findings show that socio-economic characteristics as measured in free school meal eligibility of pupils within schools are an important factor in school response. The effect questions whether the inclusion of replacement schools based on the current scheme improves the representativeness of 15 year olds in the sample. If replacement schools are used, it seems more important to match initial and replacement schools on the socio-economic composition of their pupils than on any of the factors currently used. To increase school response amongst replacement schools such schools may be approached earlier during the data collection stage to allow for a longer response period.

Pupil-level response and implications for survey practice

School response patterns differ greatly from pupil response patterns, indicating that designs of education surveys need to consider different response mechanisms at both levels for achieving the best possible representative sample. In contrast to school response, pupils' ability is the strongest predictor of pupil response throughout all models, while for example pupils' free school meal eligibility does not gain any importance conditional on ability. Results of the multilevel models show that schools matter for pupil-level response since the between school variance is highly significant and 15 per cent of variation in the pupils' response propensity is due to differences between schools. Examining contextual effects with the aim of explaining such school-level influences, we do not find that pupils in private schools, schools with high achievement pupils and schools with low free school meal eligibility show different response behaviour from their peers in other schools. Although we have rich data available, it is notable that our school variables are limited to providing information on pupil populations within 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. It is therefore recommended that future designs collect more information that may be related to the non-response process, including paradata (such as time stamps (e.g. time of survey requests, time between requests, time until response/non-response), reasons for non-response etc).

Given the importance of achievement for pupil response, we tested whether the association between pupil 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 the response probability. It follows that schools with a higher pupil-level response rate may contribute more to the risk of non-response bias than schools with a lower pupil-level response rate. This implies that increases in response rates may not lead to a reduction in the non-response bias and illustrates the potential danger of using the response rate as a data quality indicator (see Groves and Peytcheva, 2008, for a discussion on the relationship between non-response rate and bias for household surveys; see Wagner (2012) and Merkle and Edelman (2009) for an example of increased response rate with increases in bias in exit polls). This means that existing practices to exclude schools with low pupil response from the dataset should be revisited. In particular, it needs to be recognised that schools with high response rates can even increase the risk of non-response bias. This is a key finding that goes against current practices in education data, where schools with high response rates are regarded as providing high quality data. It is therefore important to analyse the representativeness based on the fully observed variables.

In more recent PISA surveys school non-response remains an important issue, although the school response rate in England has increased slightly since then (in 2012 it was 77.6 per cent before replacement (Department for Education, 2014), which compares to 73.4 per cent for 2003 data), following the introduction of a school incentive to encourage participation (OECD, 2012). Also since 2006, the England PISA test has taken place about 6 months before the GCSE exams, which reduced the survey burden of replacement schools to a certain extent.

To better understand the non-response process in future surveys (PISA or education surveys more generally) we recommend carrying out a non-response analysis of a similar nature as done here (details will depend on the type of fully-observed variables available). We recommend collecting more paradata, as is already done for household surveys, which would enrich the type of variables available for non-response analysis and would help predict non-response of schools and pupils (see for household surveys Durrant *et al.* (2015)). Replacement of non-responding schools, which generates one example of paradata, is often used to boost response in education surveys. However, given the results from this analysis, non-response of such schools is lower than for the original sample units even after controlling for other auxiliary variables. If part of the design includes the use of replacement schools, careful consideration is also necessary of how best to select such replacement schools. Any stratification that leads to the selection of replacement schools should be related to factors associated with higher non-response (in our case FSM eligibility). It may also be possible for survey designers to carry out a pilot in advance of the main stage survey to explore the factors that relate to non-response and then subsequently to stratify on characteristics that are associated with lower response, rather than to set these criteria a priori as is current practice. We found that the school level impacts significantly on pupil level response, which implies that schools indeed have a significant influence on whether their pupils respond or not. Hence, we recommend working closely with schools to encourage them to boost pupil level response. We also recommend collection of more information that could be related to the non-response process such as 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 schools' previous response behaviour. A key finding here, of importance to any education data, is that schools with a higher pupil-level response rate contribute more to the risk of non-response bias than schools with a lower pupil-level response rate. Hence, increasing the pupil response rate will not necessarily lead to a reduction in non-response bias and could even increase it. We recommend not using the response rate as a single indicator of data quality. Practices in education surveys to exclude schools with low pupil response from the dataset should be revisited.

Future research aimed at evaluating the quality of educational achievement surveys needs to investigate whether non-response patterns found for England can be generalised over time and across countries.

Acknowledgements:
We would like to thank Prof John Micklewright for earlier contributions to this paper as well as the Editor, Associate Editor and three anonymous referees for their very helpful comments. This research was funded by the ESRC grant Administrative Data Research Centre for England, grant number: ES/L007517/1, and by workpackage 1 of the ESRC grant The National Centre for Research Methods, grant number: ES/L008351/1. It builds on earlier work funded by the (then) Department for Education and Skills (DfES). We thank Cristina Bratu for research assistance and staff at DfES, the Office for National Statistics, and the Fischer Family Trust for their help in setting up the data. We alone remain responsible for their use.

References

Adams, R.J. (2003). Response to 'Cautions on OECD's Recent Educational Survey (PISA)', *Oxford Review of Education*, 29, 3, 377-389.

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

Araujo, L., Saltelli, A., and Schnepf, S.V. (2017) Do PISA data justify PISA-based education policy? *International Journal of Comparative Education and Development*, 19, 1: 20-34.

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.

Couper, M. (1998). Measuring survey quality in a CASIC environment. *Proceedings of the Survey Research Methods Section*, ASA, pp. 41-49.

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.

Department for Education (2014) Achievement of 15-Year-Olds in England: PISA 2012 National Report. Available from https://www.gov.uk/government/publications.

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.

Durrant, G.B., Maslovskaya, O., Smith, P.W. (2015) Modelling Final Outcome and Length of Call to Improve Efficiency in Call Scheduling, *Journal of Survey Statistics and Methodology*, 3, 3, 397-424.

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 and 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.

Groves, R.M. and Peytcheva, E. (2008) The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis, *Public Opinion Quarterly*, 72, 2, 167-189.

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

Huber, P. J. (1967) The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, Vol. 1, 221–233.

Kreuter, F. (2013) *Improving Surveys with Paradata, Analytic Uses of Process Information*, New York, Wiley.

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.

Lepkowski, J. M. and Couper, M. P. (2002) Nonresponse in longitudinal household surveys, in: *Survey Nonresponse*, Eds. R.M. Groves, D.A. Dillman, J.L. Eltinge & R.J.A. Little, pp. 259–272. New York: Wiley.

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

Merkle, D.M. and Edelman, M. (2002) Nonresponse in Exit Polls: A Comprehensive Analysis, in: *Survey Nonresponse*, edited by Groves, R.M., New York, John Wiley and Sons, 243-257.

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.

Neuhaus, J. M., Kalbfleisch, J. D., and Hauck, W. W. (1991) A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data, *International Statistical Review,* 59, 25-35.

Olson, K., and Groves, R. M. (2012) An Examination of Within-Person Variation in Response Propensity over the Data Collection Field Period, *Journal of Official Statistics*, 28, 29-51.

Olson, K., Smyth, J. D. and Wood, H. M. (2012) Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination, *Public Opinion Quarterly*, 76, 611 – 635.

Organisation for Economic Co-operation and Development (2002) PISA 2000 Technical Report, edited by Ray Adams and Margaret Wu, Programme for International Student Assessment, Organisation for Economic Co-operation and Development. Available from: http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33688233.pdf

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 (OECD).

Organisation for Economic Co-operation and Development (2005) PISA 2003 Technical Report, Programme for International Student Assessment, Organisation for Economic Co-operation and Development, Paris: OECD Publishing. Available from: http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/35188570.pdf

Organisation for Economic Co-operation and Development (2012). PISA 2009 Technical Report. Paris: OECD Publishing. Available from http://www.oecd.org/pisa/pisaproducts/50036771.pdf

Organisation for Economic Co-operation and Development (2014) PISA 2012 Technical Report, edited by Ray Adams and Margaret Wu, Programme for International Student Assessment, Organisation for Economic Co-operation and Development, Paris: OECD Publishing. Available from: http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

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.

Plewis, I., Ketende, S., and Calderwood, L. (2012) Assessing the Accuracy of Response Propensities in Longitudinal Studies, *Survey Methodology*, 38, 2, 167-171.

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.

Prais, S.J. (2003). Cautions on OECD's Recent Educational Survey (PISA), *Oxford Review of Education,* 29, 2, 139-163.

Rabe-Hesketh, S. and Skrondal, A. (2012a) *Multilevel and Longitudinal Modeling Using Stata*, Vol. I and Vol. II, 3rd ed., Texas, Stata Press.

Rabe-Hesketh, S. and Skrondal, A. (2012b) *Multilevel and Longitudinal Modelling using Stata*, Vol. II, Categorical responses, counts, and survival. 3rd ed., Stata Press, Texas.

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.

rates in school surveys. *Research Report 721*. Department for Education and Skills, London.

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.

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*, New York, Springer.

Wagner, J. (2012) Research Synthesis: A Comparison of Alternative Indicators for the Risk of Nonresponse Bias, *Public Opinion Quarterly*, **76**(3), 555-575.

West, B. T., and Groves, R. M. (2013) A Propensity-Adjusted Interviewer Performance Indicator, *Public Opinion Quarterly,* 77, 352-74.

White, H. (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817–830.

White, H. (1984) Asymptotic Theory for Econometricians, Orlando, FL: Academic Press.

White, H. (1994) Estimation, Inference and Specification Analysis, New York: Cambridge University Press.

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 across the two years.

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 across the two years.

Table 3. Estimated coefficients in five 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 ps² / 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

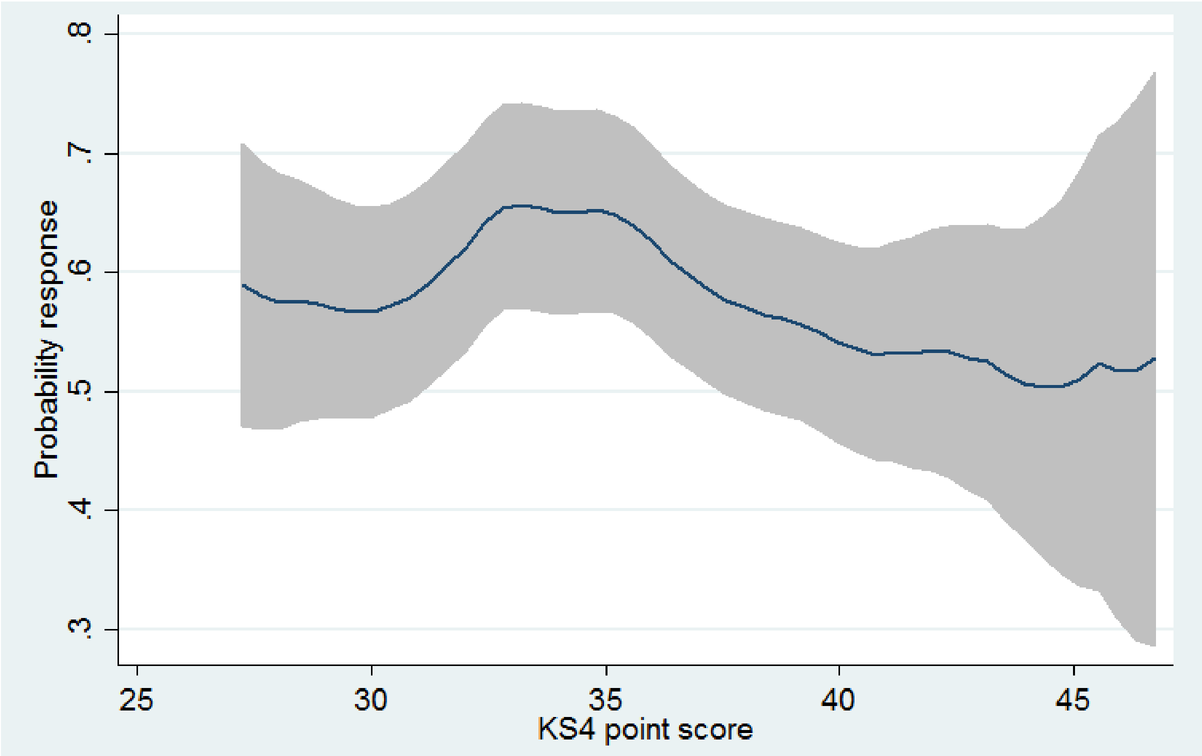
Note: school sample size is 509, standard errors are reported in parentheses,
FSM = free school meal, KS4 ps = Key Stage 4 point score
*** denotes p<0.01, ** p<0.05, * p<0.1.

Table 4. Estimated coefficients of multilevel logistic models for pupil response

				Random intercept models					Random intercept and slope models	
		(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Student characteristics	Ln(Ks4 ps), centred				0.776***		0.782***	1.028***	0.827***	1.164***
					(0.042)		(0.042)	(0.243)	(0.052)	(0.290)
	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	0.482		0.558
						(0.313)	(0.314)	(0.316)		(0.335)
	Proportion FSM in school					0.270	-0.118	-0.041		-0.179
						(0.645)	(0.664)	(0.668)		(0.668)
	Missing FSM					-0.930	-0.586	-0.586		-0.659
						(0.582)	(0.593)	(0.593)		(0.649)
	Proportion male in school					0.135	-0.220	-0.222		-0.216
						(0.328)	(0.336)	(0.336)		(0.350)
	Mean Ks4 ps, centred					0.0236	-0.0134	-0.012		-0.015
						(0.013)	(0.013)	(0.013)		(0.014)
	Year 2000			0.499***	0.490***		0.485***	0.482***	0.493***	0.490***
				(0.109)	(0.109)		(0.112)	(0.112)	(0.110)	(0.113)
	Mean pupil x mean school Ks4 ps /100, centred							-0.745		-0.974
								(0.721)		(0.827)
	Constant	1.199***	1.347***	-1.290***	1.040***	1.230***	1.143***	1.580***	1.055***	1.715***
		(0.025)	(0.056)	(0.088)	(0.084)	(0.203)	(0.205)	(0.609)	(0.0853)	(0.623)
Random school effects	VPC		0.159	0.153	0.152	0.151	0.150	0.150		
	s_{u0}		0.790***	0.770***	0.768***	0.764***	0.763***	0.762***		
			(0.047)	(0.047)	(0.047)	(0.046)	(0.047)	(0.047)		
	s^2_{u0}								0.613***	0.604***
									(0.076)	(0.075)
	s^2_{u1}								0.135**	0.139**
									(0.051)	(0.052)
	s_{u01}								0.109**	0.108**
									(0.047)	(0.047)
	Log-likelihood	-4805.8	-4557.2	-4337.1	-4350.1	-4550.3	-4348.7	-4348.1	-4341.1	-4338.8

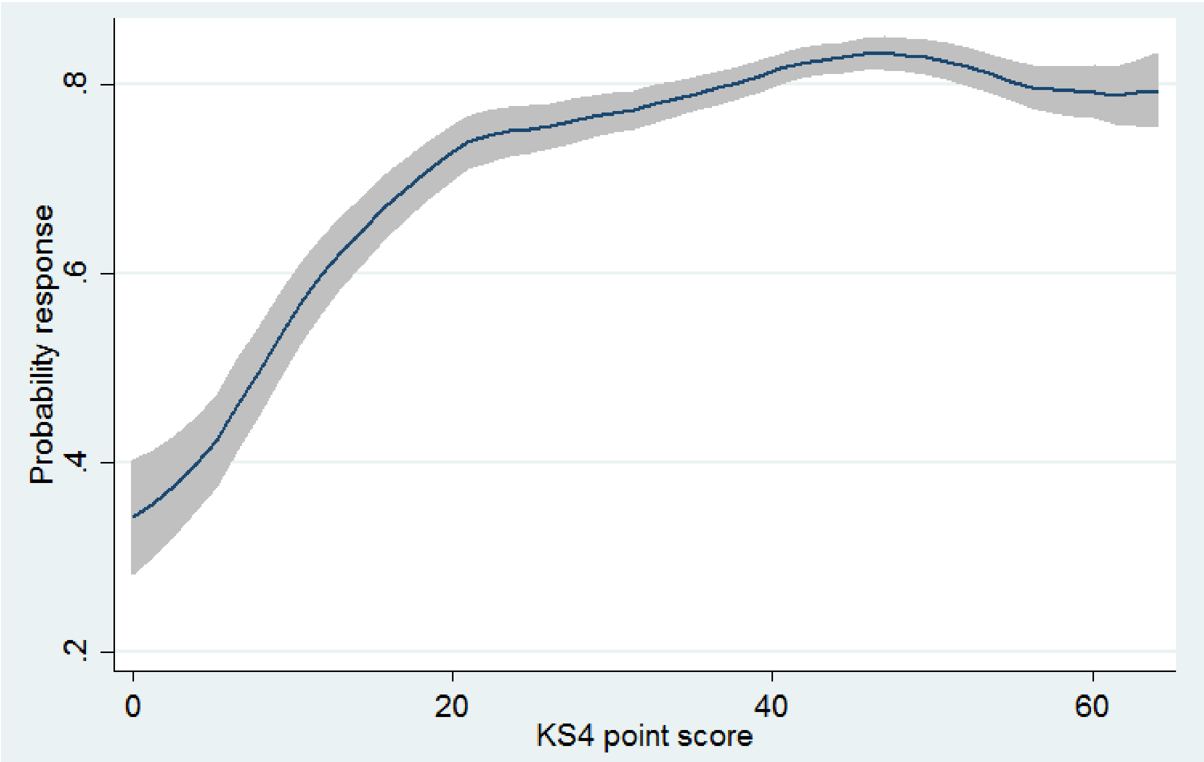
Note: results are based on 8,878 pupils in 279 schools. Standard errors in parentheses;
KS4 ps = Key Stage 4 point score, FSM = free school meal
***denotes p<0.01, **p<0.01, *p<0.05.

Figure 1. 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: The graph provides a local polynomial smooth of the achievement variable KS4 points on school response. The grey shading provides the 95 percent confidence interval. Average achievement score values between 10th and 90th percentile are taken into account.

Figure 2. Response probability of pupils by pupils' total Key Stage point score, for pooled 2000 and 2003 data using non-parametric local polynomial regression



Note: The graph provides a local polynomial smooth of the achievement variable KS4 points on pupil response. The grey shading provides the 95 percent confidence interval.