

A Study of Suitable Methods for Raising Response Rates in School Surveys

Patrick Sturgis¹, Patten Smith² and Graham Hughes¹
University of Surrey¹, BMRB Limited²

Research Report
No 721

*A Study of Suitable Methods for Raising
Response Rates in School Surveys*

*Patrick Sturgis¹, Patten Smith² and Graham Hughes¹
University of Surrey¹, BMRB Limited²*

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education and Skills.

© BMRB International Limited 2006
ISBN 1 84478 686 2

Table of Contents

1	Executive Summary	1
2	Introduction and Background to the Project.....	6
3	Survey nonresponse and school surveys	9
3.1	Response Rates in Establishment surveys	11
3.2	Response Rates and Nonresponse Bias	14
4	Trends in UK School Response Rates over Time	18
5	Implementation of PISA 2003	23
5.1	What Went Wrong?	23
5.1.1	Continuity of Project Management.....	23
5.1.2	Clarity concerning the response rate target.....	24
5.1.3	A Clear 'A Priori' Contact and Persuasion Strategy	25
6	PISA Response Rate Calculations	27
6.1	Response rate Benchmark Criteria	27
6.2	The Use of Replacement Sampling.....	31
6.3	Response Rate Thresholds	37
7	Qualitative Interviews with Nonresponding Schools to PISA 2003	43
7.1	Methodology	43
7.2	Summary of Key Findings	43
7.3	General feelings about School-based research	46
7.4	Relevance and Benefit	49
7.5	Detracting from Core Priorities	50
7.6	Monetary incentives	51
7.7	Provision of feedback	52
7.8	Repetition and Redundancy of Requests	54
7.9	Timing in the School Year	55
7.10	Teacher Workload	57
7.11	Awareness of PISA	59
7.12	Views on international comparative school surveys	60
7.13	Views on PISA fieldwork requirements.....	61
8	Survey of School Experiences and Attitudes to Survey Research	65
8.1	Rationale and Sample Design	65
8.2	Results	65
9	Recommendations	75
10	References	80
11	Appendix	1

1 EXECUTIVE SUMMARY

Anecdotal evidence suggests that surveys of school pupils in England have been in steady decline over recent years. The problem of declining response rates to school surveys was brought into sharp focus in 2003 by the failure of England to meet international response rate benchmarks for the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). Failure to meet benchmark response rate requirements for these comparative surveys had serious consequences, resulting in the England's exclusion from published tables of international student achievement, despite the large government investment in the data collection activities. It is not clear what underlies the apparent downward trend in school response rates, nor why England has consistently experienced greater difficulty than the majority of its international partners, including Scotland and Northern Ireland, in maintaining response rates and meeting the benchmark criteria.

In March 2005, the Department for Education and Skills (DfES) commissioned BMRB International Limited and Dr Patrick Sturgis to conduct a programme of research into the factors underlying the increasing difficulty researchers face in achieving satisfactory response rates in surveys of schools. This report sets out the findings of this research programme. The report has three primary foci; an investigation of trends and causes of nonresponse to UK school surveys in general; a specific focus on the causes of nonresponse to the PISA survey in its most recent round, 2003; and making recommendations for improving response rates to school surveys in England in the future. These are addressed through a combination of desk research and the collection and analysis of both qualitative and quantitative data. The key findings of the report are summarised in bullet form below.

Response rates to DfES school surveys in the UK are in decline. Our analysis of 73 separate surveys conducted by DfES between 1995 and 2004 shows an average decline of approximately 2% per annum over this period. This confirms anecdotal evidence from data collection agencies and survey sponsors as well as trends observed for a number of repeated cross-sectional surveys.

There is almost no published literature specifically addressing response rates for surveys of schools. Encouragingly, the published literature on response rates for surveys of organizations, of which school surveys are a subset, shows that standard response rate enhancement measures, such as the provision of incentives, the reduction of survey burden, and coordination of sampling strategies across the target population can be effective in raising response rates. The existing literature on surveys of organizations is based on surveys of businesses rather than schools. Care must, therefore, be taken in generalising the conclusions from these publications to the context of school surveys.

The use of arbitrary 'large numbers' as response rate benchmark criteria are of dubious utility for assessing survey quality. It is perfectly possible for a survey meeting such criteria to produce more biased estimates than a different survey which fails to meet the criteria. More emphasis should be placed, where possible, on bias assessments which compare responders to nonresponders on key survey variables. Where such comparisons indicate the presence of nonresponse bias, it would be sensible to implement weighting adjustments, rather than discarding the data on the basis of potential bias in the unadjusted data.

The use of replacement sampling as a means of reducing non-response bias, as implemented in the PISA survey, is of questionable and usually untestable validity.

Given the non-equivalence of fieldwork characteristics between the initial and replacement samples, it is possible that the practice of replacement sampling increases rather than reduces biases in the initial sample. Calculations which insert 'substituted' sample units into the numerator but not the denominator of the response rate equation are misleading as an indicator of potential nonresponse bias. It would be relatively straightforward to investigate the effectiveness of replacement schools in reducing bias in estimates of pupil achievement using the UK data from PISA 2003. Because bias can be reliably estimated via weighting adjustments using GCSE scores of responding and nonresponding pupils, the effect of replacement schools on the degree of bias could be estimated by differencing the bias at first and second phase of sampling⁴.

The PISA technical standards implement a formula which increases the post-replacement response rate benchmark by a half percent, for each percent the first phase rate falls below the first phase benchmark of 85%. If we make the joint assumptions that the first phase rate will be unlikely to reach 85% and that the second phase rate will be equal to the first phase rate, then the minimum first phase response rate needed to meet the post-replacement benchmark is 72% if only first substitute schools are issued. First phase response rates below 68% would render use of the replacement sample a potential waste of public funds as well as school and pupil time. Assuming that both first *and* second substitute schools will be used and a drop in response rate of 5% at each phase of fieldwork renders a target response rate of only 68% for the initial sample.

Implementation of PISA 2003 in the UK suffered from a number of organisational and administrative flaws, from which lessons should be drawn for the 2006 round of the survey. First, measures should be taken to keep the turnover of key staff in the data collection agency to a minimum. Second, a revised first phase response

rate benchmark (see above) should be employed, which is agreed by all stakeholders as achievable from the outset. Third, a clear, adequately resourced, 'a priori' contact and persuasion strategy should be agreed at the outset and stuck to.

Fieldwork for PISA 2006 should be moved to the autumn. While interviews with head teachers indicated that, with regard to surveys in general, no time of year is viewed as optimal, implementing PISA in the spring is particularly problematic for 15 year old pupils as they are preoccupied with preparation for their imminent GCSE examinations. Additionally, for a spring fieldwork period, if it is necessary to issue the replacement sample, many sampled pupils will have *de facto* left school by the time contact is made. Conducting fieldwork in the autumn would avoid this potentiality.

Qualitative interviews with Head teachers in schools which declined to participate in PISA 2003 indicate that schools are now receiving very large numbers of requests to participate in research. In deciding whether to agree to such requests, Head teachers mainly focus on the relevance and benefit of participating to their particular school. In general, they do not give any particular weight to survey requests simply because they come from DfES. When they decide not to participate, it is mainly due to insufficient time, the distraction from a school's core activities, the perceived repetition and redundancy of requests for information, and the perception that schools and pupils are already excessively audited.

A small probability survey of schools in England and Wales confirms the general picture to emerge from the qualitative interviews; schools feel 'over-surveyed', believe that educational research is poorly coordinated, is often of no relevance or

benefit to pupils, and requires scarce time and resource. A clear majority of schools currently rate their probability of participating in a PISA-type survey in 2006, if requested, as either 'fairly' or 'very unlikely'. Encouragingly, however, schools consider that changes to the way in which fieldwork is implemented in future rounds would make them more likely take part.

2 INTRODUCTION AND BACKGROUND TO THE PROJECT

This report sets out the results of a programme of research investigating the causes of apparent declines in response rates to the Department for Education and Skills' (DfES) portfolio of domestic and international school surveys over the past five to ten years. Despite the implementation of a range of response boosting strategies, surveys of pupils in England appear to have been in steady decline and have failed to meet international response rate benchmarks, most prominently for the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) survey. Although potential causes of this decline abound, It is not fully clear what underlies this apparent trend, nor why England has experienced greater difficulty than the majority of its international partners in maintaining response rates and meeting the benchmark criteria. Failure to meet benchmark response rate requirements for these comparative surveys has serious consequences, potentially resulting in their exclusion from published tables of international student achievement, despite the large government investment in the data collection activities.

The report has three primary foci; an investigation of trends and causes of nonresponse to UK school surveys in general; and a specific focus on the causes of nonresponse to the PISA survey in its most recent round, 2003; and making recommendations for improving response rates to school surveys in England in the future. We address these primary research questions through a combination of desk research and the collection and analysis of both qualitative and quantitative data. The remainder of this report is divided into seven overlapping sections.

We begin by reviewing the published literature on survey nonresponse, with a particular focus on surveys of establishments, of which school surveys are a part. In this section we assess the primary explanations for declining response rates to this type of survey and examine in detail the relationship between survey nonresponse and estimation bias.

Next we proceed to a systematic and empirically based evaluation of the idea that response rates to school surveys in the UK have been in general decline for the past five to ten years. Much of the case supporting the thesis of decline is anecdotally based, or founded on response rates to single studies, repeated over a relatively brief time frame. Using the online DfES publications database as our starting point, we compile and analyse response rates to DfES school surveys between 1995 and 2004.

Having established a more robust descriptive picture of the trend in response rates to school surveys in the UK in recent years, we proceed to a detailed description of the way fieldwork for PISA 2003 was implemented. We base this account on unstructured interviews with key personnel from the data collection agency, the Office for National Statistics, and the commissioning department, the Department for Education and Skills. After presenting a clear picture of what actually happened 'on the ground', we proceed to a critical analysis of what went wrong and draw out implications for the conduct of the next round of the survey in 2006.

In section six of the report we focus on the rationale underlying the response rate calculations and benchmark criteria employed in the PISA survey. In particular, we examine the implications of using a weighted rather than an unweighted response rate and the use of substitution from a replacement sample of schools. We

introduce a way of determining the likely 'post-replacement' response rate on the basis of the first phase response rate and describe how this might be of use in achieving benchmark response rate criteria and preventing the potential waste of public funds on needless second phase data collection.

In section seven, we report on twenty five qualitative interviews conducted with head teachers from schools which declined to participate in PISA 2003. Our analysis focuses on the factors and survey characteristics which heads see as important in influencing their decision of whether or not to participate in survey research. Initially, we focus on requests for survey participation in general, before moving on to consider the specific case of PISA in more detail.

Section eight of the report contains analysis and interpretation of a small quantitative survey of head teachers in England. Building on the findings from the qualitative interviews, the survey provides us with a clearer idea of the extent and distribution of perceptions and experiences of survey research within schools across the country. In particular, the survey provides interesting insights into the frequency with which schools are now receiving requests for survey participation and how this has been changing over time. Other questions ask Heads to detail what makes them more or less likely to agree to survey requests, particularly in the context of international surveys of student achievement like PISA.

We finish by drawing together all the various strands of the report to produce a set of actionable recommendations for boosting response to school surveys, with particular reference to PISA 2006.

3 SURVEY NONRESPONSE AND SCHOOL SURVEYS

A serious hazard to drawing inferences from sample surveys is that a substantial proportion of the random sample of the population originally selected may fail to respond. Quite apart from the loss of sample size which this entails, the non-responding units are likely to differ from responding units in terms of characteristics which are directly or indirectly related to the variables which the survey is intended to measure. Where this is so, nonresponse can seriously bias univariate and associational population estimates.

The literature on household surveys contains many instances of demonstrated or inferred nonresponse bias of this kind. Non-responders have often been shown to differ from responders in terms of a number of socio-demographic and economic variables which are linked to lifestyles, attitudes and beliefs (Lievesley, 1983; Groves and Couper 1998). The pattern of nonresponse bias may be complex, because the characteristics of those who cannot be contacted can differ markedly from the characteristics of those who are contacted, but refuse to participate (Wilcox, 1977).

There is, furthermore, evidence that response rates to general population surveys have been in decline over recent decades (cf Steeth, 1981; Lievesley, 1983; Goyder, 1987; Bradburn, 1992, de Leuw and de Heer, 2002). Trends in living patterns and attitudes have increased the problems faced by research agencies in obtaining acceptable levels of response in two main ways. It has become more difficult to find people at home (Scott, 1971); and it has become more difficult to persuade people, once contacted, to take part in interview surveys (Steeh, 1981). This general decline in cooperation has been linked to long-term processes of societal atomisation, increasing personal contact from survey, market research

and direct marketing organisations and a general decline in social capital (Schleiffer, 1988; Groves and Couper, 1998; Campanelli et al 1998). Some government continuous surveys in both the USA (Groves, 1989) and Great Britain have maintained their overall rates of response, but even in these cases refusal rates have tended to rise, compensated by lower non-contact rates achieved at higher cost in time and money than many other surveys can afford (Groves, et al 2004).

It is important to note that this evidence relates primarily to surveys of households and individuals. The focus of this programme of research, however, is on surveys of schools, which, are different in important ways from household surveys. Although, we have been unable to identify any published work in the specific area of nonresponse in school surveys, we can draw on a number of publications in the more general area of surveys of organizations, of which school surveys are a subset (Groves et al 2004). However, the majority of published work in this area is itself based on surveys of businesses and establishments, which are not directly comparable to schools as organisational entities. While we believe that this body of literature is of utility in understanding nonresponse in school surveys, the indirect nature of the comparison should be borne in mind.

The evidence for a general decline in response rates on establishment surveys, is not so strong as in household surveys (Atrostick et al 1999). Although there is a perception amongst survey managers in international statistical agencies that it is becoming harder to maintain levels of response with the same unit resource (Christianson and Tortora, 1995). Establishment surveys are also rather different from general population surveys, in that the categories of nonresponse tend to be fewer, cooperation is more often mandatory, and interview mode is generally self-

completion as opposed to interviewer administered (Paxson, Dillman and Tarnai, 1995). Nonetheless, there is good reason to believe, both anecdotally and on the basis of response rates to a number of repeated cross-sectional surveys, that response rates to school surveys in the UK are in a similar state of decline.

The perceived drop-off in response rates to school surveys over the past five to ten years coincides with the implementation of a number of important changes in the UK compulsory education system. These centre on an increased administrative burden on schools and teaching staff through the introduction of key skills tests for seven, eleven and fourteen year olds, greater weight afforded to coursework based assessment and external audit through the Office for Standards in Education (OFSTED). The report produced by the US NISS/ESSI Task Force on Participation Rates in International Assessments concluded that similar forces underlie the increasing reluctance of schools to participate in the PIRLS, PISA and TIMSS surveys in the US (NISS, 2004) Thus, if response rates to school surveys are in decline, the trend is readily explicable in the confluence of these factors serving to reduce both the time teachers have to participate in research studies and their general willingness to complete paperwork and administer tests. Later, in section 4 of this report, we evaluate empirically the case for a general decline in response rates to school surveys in the UK over the past ten years.

3.1 Response Rates in Establishment surveys

As with household surveys, low response rates in establishment surveys are assumed to be strongly linked to the general survey environment. This is predominantly outside of the survey manager's control and includes factors such as the social structural trends and changes in wider society noted earlier, in addition to short to mid term systemic 'shocks', such as war and political scandal

(Groves and Couper 1998). Encouragingly, however, research into the correlates and causes of nonresponse in establishment surveys shows that response rates can be improved through the implementation of procedures commonly used in household surveys (Petroni et al, 2004). Thus, Christianson and Tortora (1995) report increases in response rates to business surveys through the reduction of survey burden, careful targeting of field resources on pre-notification and persuasion of sampled units, and focused nonresponse follow-up.

Similarly, monetary incentives have been shown to be effective in boosting response in establishment surveys (Jobber, 1986; James and Bolstein 1992). This runs counter to the findings of a study into the effects of incentives on nonresponse to PISA and TIMSS in the US, where such incentives were seen as irrelevant to the participation decision amongst nonresponding schools (Westat, 2003). It is important to note, however, that this latter study suffers from a serious limitation, in that it only collected data from nonresponding schools and may, therefore, simply reflect *post hoc* rationalisation and socially desirable responding amongst a non-random subset of schools. Paxson, Dillman and Tarnai (1995) demonstrate the effectiveness of telephone follow-ups to initial nonresponders and find strong evidence of higher response rates when an establishment survey becomes mandatory. Mandatory surveys were also found to elicit significantly higher response rates in surveys of businesses by Willimack et al (2002). An informal survey of international experiences in gaining cooperation to PISA conducted by DfES in 2004 reveals that PISA is either mandatory or 'quasi-mandatory' in many of the countries that achieve high response rates to the survey.

A more holistic approach to institutional data requirements can also pay dividends through rationalisation and coordination of survey requests across the population of establishments; Ohlsson (1995) and Siranth and Carpenter (1995) find higher response rates to Census Bureau establishment surveys after implementing a rotational sample design, which minimised the amount of overlap in survey requests received by the same establishment within a fixed period.

Lastly, it is important to note that the extant literature on maximising response rates in establishment surveys places a great deal of weight on contact and persuasion with the establishment 'gatekeeper' (Cox et al 1995; Christianson and Tartora 1995; Willimak et al 2002). Most of the time, initial contact will need to be made with an employee who has both the power to authorise institutional participation and access to the requisite information. Such individuals are usually senior, with little time available for what might be considered 'non-essential' activities. And, herein, lies the primary problem of achieving high response rates to establishment surveys, for non-mandatory surveys are never likely to feature very high on the 'things to do' lists of such individuals. In school surveys, of course, the 'gatekeeper' is almost always the Head teacher¹. Thus, successful strategies for boosting response rates to school surveys will focus predominantly on maximising the motivation of these key individuals to participate in the survey in question. In sections 7 and 8 of this report we employ qualitative and quantitative interviews with Head teachers in UK schools to identify the sorts of

¹ Sometimes the Head teacher may delegate such decision-making to a colleague. However, such people are likely to share the key characteristics of being senior employees and having scarce time for non-mandatory activities.

things which predispose Head teachers to accept or decline requests for survey participation made to their schools. Before then, however, we look in more detail at the issue of how survey nonresponse is related to nonresponse bias and how a simple focus on the magnitude of the response rate can be misleading as a measure of survey quality.

3.2 Response Rates and Nonresponse Bias

Survey response rates are often interpreted as an indicator of survey quality in and of themselves. However, bias in survey estimates is not a simple function of response rate, but arises from *differences between respondents and nonrespondents on relevant survey variables*. This is summarised by Groves (1989) as follows:

$$y_r = y_n + \left(\frac{nr}{n} \right) (y_r - y_{nr})$$

where n represents the total sample eligible for the survey, r the sample units which respond to the survey, nr the sample units which do not respond, and y a survey estimate (such as a mean, a proportion, or regression coefficient). Here we can see that the sample statistic (y_r) is a function of the value we would have obtained given 100 percent response (y_n) and a non-response error term. This term shows that the extent of non-response bias is a function of the size of the proportion of non-respondents (nr/n) and the magnitude of the difference in characteristics between respondents and non-respondents ($y_r - y_{nr}$). Because the latter term, $y_r - y_{nr}$, is, almost by definition, rarely available, the response rate is often treated as a measure of data quality *per se*, rather than merely an indicator of potential bias.

This is not to deny that, *ceteris paribus*, surveys with low response rates are likely to produce more biased estimates than surveys with high response rates. Indeed, common sense and empirical investigations alike suggest that nonresponse bias is usually greater on surveys with higher nonresponse rates (see Foster and Bushnell, 1994). More recently, however, researchers have begun to question the idea of a neat, deterministic correspondence between response rate and nonresponse bias (Stoop 2005; Groves et al 2004). For instance, Merkle and Edelman (2002) report response rates to cross-national exit polls for the 2000 US Presidential election varying between 10% and 90%, with the majority of polls falling within the 45-75% range. Despite this wide variation in the response rate across individual polls, they found no correlation between the level of nonresponse and the magnitude of bias in the estimate of Democratic or Republican vote percentages. Similarly, Keeter et al (2000) compare two RDD surveys that were identical but for the length of the fieldwork period and effort put into obtaining interviews. The first survey achieved a response rate of 61%, while the survey with the shorter field period and lower interviewer effort attained a response rate of just 36%. Despite these wide differences in response rate, the authors found that on only 15% of items were there significant differences between the two surveys.

Voogt (2004), too, concludes that associational relationships are usually unaffected by variation in response rate. Indeed it is possible that increasing a response rate above an initial level might increase the bias of survey estimates. For instance, increasing a response rate from 60% to 65% by targeting initially hard-to-contact respondents might serve to increase the Mean Square Error of survey estimates if the 5% of hard-to-contact respondents are highly atypical on key survey variables (Groves 1989). An example of this effect comes from a study by Moore and Tarnai (2002), who found that an incentive used in a mail survey

was particularly attractive to women and older respondents, leading to their over-representation in the final sample. In this case, the higher response rate engendered by the incentive resulted in more biased estimates on key survey variables.

Because of the complexity of the relationship between response rate and the magnitude of bias in survey estimates, few survey methodologists will be found who support 'arbitrary' response rate cut-off criteria, of the sort employed by PISA (cf Fowler 2002; Groves 1989; Platek and Sarndal 2001; Biemer and Lyberg 2003; Martin 2004). The PISA technical standards specify a minimum 85% weighted pre-replacement response rate at the school level and a minimum 80% weighted response rate at the pupil level for inclusion in the comparative tables in the published OECD report. Where a country fails to meet the pre-replacement threshold, they can make an adjusted response rate threshold through the inclusion of replacement schools. If a school fails to meet the pre and post replacement benchmarks they cannot be included in the published tables if the pre-replacement response rate is below 65% (see section 6 for a detailed discussion of PISA response rate calculations and benchmark criteria).

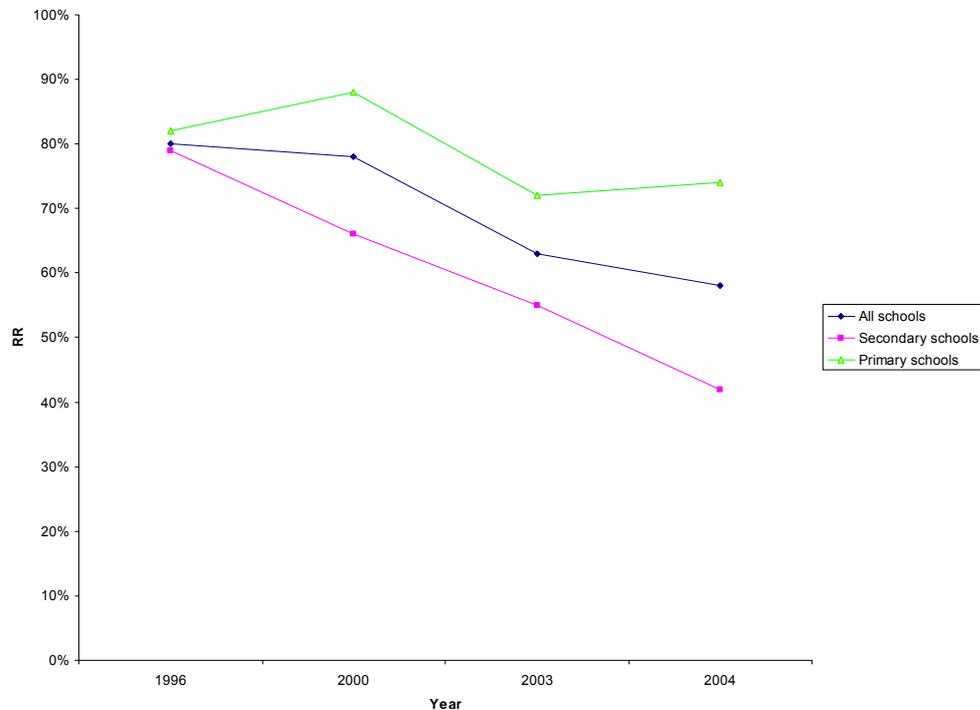
Although all survey researchers may agree that obtaining a high response rate is generally 'a good thing', the key problem with this kind of criterion based approach is that a survey failing to meet the response rate target may yield less biased estimates than another survey which easily exceeds it. As Betsy Martin noted in her 2004 Presidential Address to the American Association of Public Opinion Research, "There does not seem to be a basis for fixing a level or range below which we can say results are compromised by nonresponse bias, and above which they aren't". (2004, p441). Thus, in evaluating the quality of survey

estimates, a bias assessment based on a comparison between responders and nonresponders on key survey variables should be considered of considerably greater worth than one which requires meeting an arbitrary, pre-defined 'large number'. This is particularly so when the bias assessment is subsequently used to make weighting adjustments; an estimate adjusted by powerful auxiliary variables should generally be preferred to an unadjusted estimate from a survey with a higher overall response rate.

4 TRENDS IN UK SCHOOL RESPONSE RATES OVER TIME

As we have noted elsewhere in this report, there is a general consensus that response rates to school surveys have been in steady decline for the past five to ten years. This is in line with trends in survey research more generally. However, much of the evidence pointing to declining response rates is either of an anecdotal nature, or is based on two or three administrations of the same survey over a relatively brief period. The Teachers' Workload Survey (TWS) is a case in point, it has been carried out to an almost identical specification in recent years by BMRB Social Research on behalf of the Office for Manpower Economics. The survey involves (i) obtaining a face-to-face school head teacher interview and (ii) time use diaries from 14 teachers per school. Figure 1 shows school level response rates obtained on TWS between 1996 and 2004.

Figure 1 TWS Response Rates 1996-2004



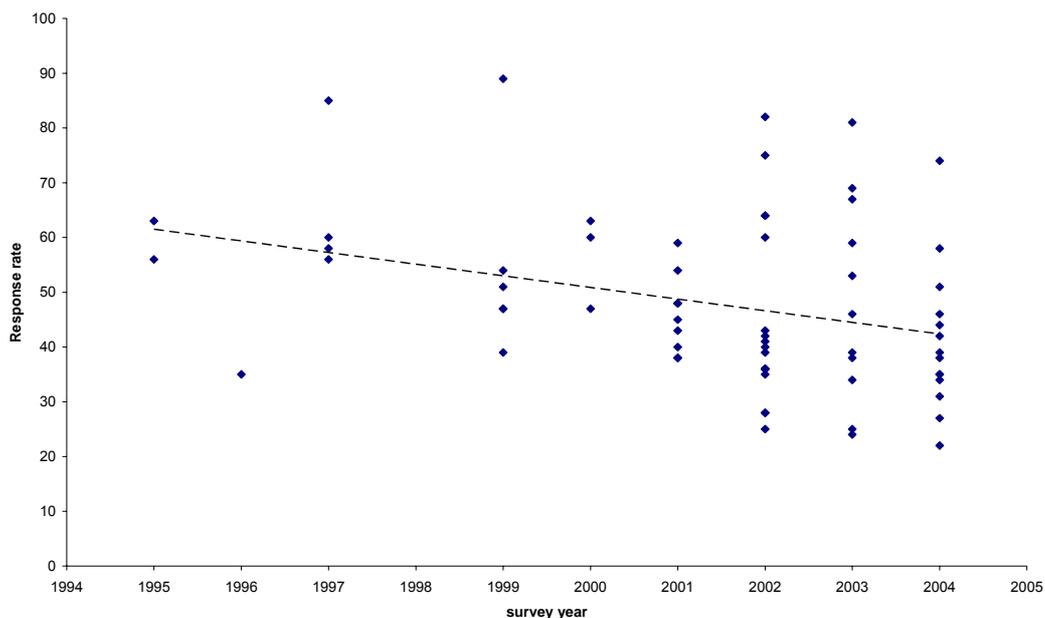
The evidence from this survey supports the hypothesis of a decline in cooperation rates in schools, with secondary schools showing a more precipitous drop-off over time than primaries. There is obviously a danger, however, in the use of single cases to support the hypothesis of a general trend. It might, for instance, be the case that only certain types of survey are affected by declining response rates but that their prominence leads to a belief that school survey response rates are in decline more generally.

At the outset of this project, therefore, we decided that it was important to establish a clear, empirically based assessment of response trends to school surveys over the past five to ten years. This proved to be no easy task, however, as no definitive register of school surveys for this period is held by DfES or any other governmental or non-governmental organisation. We recommend that the DfES begins to produce and archive meta-data on the research it conducts on its own account and through its external contractors. Such an archive would contain basic information on the aims of the study, the collectors of the data, target population, issued and achieved sample sizes and response rate. This would not be particularly onerous for those conducting school based research and would be of great benefit to DfES in keeping track of the quantity, quality and focus of the research it conducts.

Given the absence of an existing database of school survey response rates, the only option was to produce one ourselves. We did this by examining all 935 published reports on the DfES online Publications database (<http://www.dfes.gov.uk/research/>) in April 2005 and extracting the response rate data from any study that had conducted a survey of schools. We then passed this list to our DfES contacts to identify any gaps in the identified studies. This resulted in a total of 73 separate school surveys that took place between 1995 and 2004.

There are clear potential limitations to the exhaustiveness of our search and the representativeness of these surveys, in particular a bias towards DfES surveys conducted more recently that are recorded and stored on the publications data. Notwithstanding the reservations we ourselves would place on the completeness and representativeness of our sample, however, the over-time response rate trend bears out the anecdotal and single survey evidence of a general decline. Figure 2 plots survey response rates against the year in which the survey was conducted (we would have liked to employ a more fine grained metric for time but the information on the database was only accurate at they year level for all studies).

Figure 2 School Survey Response Rates 1995-2004



Although there is a good deal of variation around the line of best fit, the overall trend is clearly downward with a Pearson correlation coefficient of -0.32 ($p=0.008$). The corresponding unstandardised OLS regression coefficient for this relationship is -0.021 , indicating an average decline in response rate of approximately 2% per year. Given the relatively small number of data points, it does not make sense to

investigate variations within this overall linear trend. A limitation to this approach to modeling the data is that it treats all surveys as providing equal weight to the average or 'pooled' effect size. This is problematic because a survey with a large sample size is a more reliable estimate of schools' propensity to respond than a survey with a small sample size (Hox, 2002).

Alternatively, then we model the data as a multi-level meta-analysis in which each observation (survey) is weighted by its (transformed) issued sample size². This is done by specifying a 2 level model with effect size at level 2, and standard errors nested within each effect size at level 1. The effect sizes are assigned a fixed component, and a random element at level 2 (the level of effect size), whilst the standard errors are only assigned a random element at level 1, with variance constrained to 1. The model was estimated using MlwiN 2.0 (Rasbash et al 2004). This strategy, however, was not materially different to the OLS estimated relationship of response rate with time, yielding a pooled effect size of -.019. Having used both approaches to estimating the linear trend, however, we can have greater confidence in drawing the conclusion that response rates to school surveys have been in steady decline for at least the last ten years.

A further advantage of adopting a multi-level framework for performing this analysis is that it allows the analyst to model variation in the pooled effect size as a function of higher level characteristics, in our case, the survey. For instance, we examined differences in the trend as a function of whether surveys had been

² The standard error of the estimate is obtained by applying the following transformation to the issued sample size $n \sqrt{1/(n-1)}$ (Lipsey and Wilson, 2001).

conducted on primary or secondary schools and the mode of survey administration. None of these factors, however, proved significant although the failure to obtain significant estimates may well result from the lack of statistical power, given the small number of observations in our sample.

5 IMPLEMENTATION OF PISA 2003

A central aim of this research is to identify factors which will improve the UKs achieved response rate in the 2006 round of PISA. Key to our ability to make appropriate actionable recommendations is the development of a clear picture of what went wrong in the previous round of the survey, where response rate benchmarks were missed by some margin and the UK data was deemed inadmissible for OECD international comparisons³. To that end, unstructured interviews were conducted with key DfES and ONS staff to reconstruct the way that fieldwork had actually been implemented. This section of the report summarises the key inferences to be drawn from the problems experienced in the 2003 round of the survey.

5.1 What Went Wrong?

The implementation of PISA 2003 cannot, to be sure, be held up as an exemplar of how school surveys should be conducted. What lessons, though, can we learn from the 2003 experience to ensure a more satisfactory conclusion to its next round in 2006? We identify three key areas of improvement:

5.1.1 *Continuity of Project Management*

The 2003 survey clearly suffered from the high turnover of staff at ONS. Not only would the bewildering array of different signatures on letters sent to schools have been confusing for those receiving them, there are obvious detrimental operational side-effects when a project of this nature experiences the departures of five key

³ Note that PISA was not unique in failing to meet benchmark criteria in 2003. TIMSS 2003 achieved an unweighted response rate of just 39%. The problem, therefore, appears to be of a general nature.

members of staff within a twelve month period. It is likely that staff turnover contributed to the seemingly 'unplanned' and 'ad hoc' nature of the contact and persuasion strategy, to which we turn in a moment. It is, of course, difficult if not impossible to prevent staff turnover completely. However, the data collection agency tendered to conduct the 2006 fieldwork should be required to explicitly state what measures it has in place to ensure that turnover of key staff working on the survey is kept to a minimum. Additionally, a clear replacement/shadow management structure should be in place in the event of unavoidable staff departures.

5.1.2 Clarity concerning the response rate target

The fieldwork for PISA 2003 began with the data collection agency explicitly stating its belief that the response rate target was unobtainable. This is clearly undesirable in terms of the effect such a position will have on the motivation and morale of the research team and fieldwork staff alike. Believing a target is achievable should be considered as almost a prerequisite to meeting it. Our analysis of school surveys conducted in the UK over the past ten years indicates that only two out of the 74 surveys identified had response rates of 85% or above (the first phase benchmark) and these were considerably less burdensome than PISA. Perhaps, then, the belief that an 85% first phase response rate to PISA is unfeasible was accurate. However, as we discuss in section 6, if we assume from the outset that the replacement sample will be used, the true target response rate is a considerably less daunting 72%. Although care must be taken not to set sights too low (and, thereby, risk undershooting the revised target) it would seem sensible to adopt this more holistic approach and provide a target response rate which, although still challenging, can be considered eminently achievable with the deployment of sufficient resources.

5.1.3 A Clear 'A Priori' Contact and Persuasion Strategy

The account of the various approaches made to and incentives offered to schools in the 2003 round of PISA reveals increasingly desperate attempts to make up for an unsatisfactory initial contact and persuasion strategy. Initially, no monetary or other incentive was offered to schools, with the first contact consisting of a long and rather dense letter explaining the benefits of participation in ways that could be of only indirect and rather tangential benefit to the schools themselves. Thereafter, schools were offered more and more reasons to participate, including endorsements from teaching unions and DfES/OECD figureheads. Larger and larger monetary incentives were also introduced, as the need to secure the necessary number of schools became increasingly desperate. Not only does this raise ethical issues concerning the equitable treatment of schools, our qualitative interviews (section 7) reveal that the 'sliding incentive' approach led a number of head teachers to wonder how much they might end up with if they kept on holding out for more.

Clearly, then, a key lesson to draw from the 2003 round of fieldwork is that the resources that ended up being put into obtaining school cooperation once the benchmark appeared in jeopardy need to be instigated at first contact. Specifically, schools should receive a brief initial letter informing them that they have been selected to participate in PISA 2006. A substantial unconditional monetary payment should be made at this first contact stage. Schools should be informed that they are free to use this money however they choose, though it may be wise to include an illustrative list of some possible uses. We also recommend, resources permitting, that an equivalent monetary payment be offered at this stage, conditional on the school's participation. The primary aim of the first contact letter should not be to secure cooperation there and then but to arrange a short face-to-face meeting with a trained interviewer, whose role would be to explain the

purposes and requirements of PISA in more detail and secure the cooperation of the school.

6 PISA RESPONSE RATE CALCULATIONS

The PISA technical standards specify a minimum 85% weighted response rate at the school level and a minimum 80% weighted response rate at the pupil level for inclusion in the comparative tables in the published OECD report. Where a country falls short of these criteria, they may still be included in the published report, through the use of replacement schools, which ‘match’ nonresponding schools on the stratification variables used in the sample design. For each sampled school, two potential replacements are identified by additionally selecting the school immediately above and below the sampled school on the stratified sampling frame. When replacement schools are used, the post-replacement benchmark response rate increases by half a percent for each percent the first phase response rate fell below the 85% threshold. No distinction is made in the post-replacement response rate calculation between first and second replacement schools. Even when the post-replacement threshold is not met, country data can still be included in the published tables if country coordinators can persuade the OECD secretariat that the data do not contain serious biases. In making this decision, the OECD secretariat is guided by advice from the PISA Technical Advisory Group, which is chaired by the PISA sampling referee. We have been unable to locate any published account of the criteria used by the OECD secretariat to determine what constitutes an unsatisfactory level of bias, nor what standard of evidence is required to allow it to come to a judgment. In this section, we discuss the rationale for and validity of these procedures.

6.1 Response rate Benchmark Criteria

Earlier, the PISA survey’s use of pre-determined benchmark response rate criteria for determining acceptable statistical accuracy was referred to as ‘arbitrary’. This denotation relates to the fact that it is implausible, as was noted in section 3 of this

report, to assume a neat linear relationship between response rate and bias; samples meeting the benchmark criterion may perfectly well be considerably more biased than others failing to meet it (see the discussion of the relationship between response rate and bias in section 3). The term 'arbitrary' is not intended to mean that the criteria have been arrived at without any underlying rationale. Indeed, the technical report for PISA 2000 sets out an explicit rationale for both the specific response rate criteria employed and the criteria for determining an unacceptable level of bias for country surveys which fail to meet the pre and post replacement benchmarks. We assume that these same criteria were applied to the 2003 round of the survey, although this is not explicitly stated in Chapter 15 on adjudication in the 2003 Technical Report.

The response rate benchmark criteria are based on projected biases under a range of different values of the correlation between response propensity and achievement, across different possible response rates. Assuming a true mean of 500 for achievement, a standard deviation of 100 and a standard error of 3.5, it is possible to forecast at what point the projected bias pushes the expected value of the mean outside the 95% confidence interval of the true population mean. On this basis, response rates of 85% will still provide an expected mean value within the 95% confidence interval when correlations between response propensity and achievement are within the range -0.19-0.19. However, the explicit statement of the rationale for the 85% response rate only begs the question of how the 0.19 upper limit for the correlation between response propensity and achievement was alighted upon, particularly as there is no obvious way of knowing what the true value of this correlation is, what a 'typical' value might be, nor whether the relationship should be expected to remain constant across countries.

The criteria for determining what is an unacceptable degree of bias for countries failing to meet the pre and post-replacement benchmarks is a good deal less precisely specified; the report states that logistic regression models are used to estimate nonresponse propensity from a range of school level predictors and that nonresponse bias may be a significant problem if any of these coefficients is “substantial”. No definition of what constitutes a substantial magnitude is provided. Furthermore, this approach only identifies whether nonresponse is differential, it says nothing about whether the groups with different response propensities also differ on the key survey variables. As can be seen from the formula on page 14, these logit coefficients could be of substantial magnitude without introducing any bias at all into survey estimates.

Where a country fails to meet the pre and post response rate benchmarks (that is less than 85% at first phase and less than or equal to $y\%$, where $y=(255-x)/2$ after replacement, that country’s data can still go into the published tables if they are able to persuade the PISA secretariat that serious biases are not present in the data⁴. The PISA 2003 Technical Report states that the “aim of the adjudication process is to make a single determination on adjudicated data in a manner that is transparent, based on evidence and which is defensible” (2005. p237). However, given the lack of published criteria detailing how the secretariat comes to a judgement on what is an acceptable or unacceptable degree of bias, it is difficult

⁴ Technically, the PISA sampling rules also state that countries achieving less than 65% pre-replacement response rates cannot be considered for inclusion under the bias analysis procedure. However, both the UK and US were considered under this procedure, despite missing the 65% target, with the US included in the published tables.

to see how the adjudication process can meet these stated aims. No such criteria appear in the Technical Standards for the Implementation of PISA 2003 document, nor in the PISA 2000 or 2003 Technical Reports and we have been unable to locate any document containing such criteria anywhere on the PISA website, or elsewhere.

The lack of explicit criteria with regard to the adjudication process can result in apparently anomalous decisions. For example, both the UK and the US failed to meet the pre and post school level response rate benchmark criteria in the 2003 round of PISA. Additionally, the UK failed to meet the pupil level response rate benchmark of 80%⁵. Although the UK was excluded from the published tables on the grounds that, “student non-response was likely to have induced a bias in achievement” (p246), the adjudication also concluded that “It was not possible to ascertain the exact magnitude of this (bias)” (p246). On the other hand, the US was included in the published tables on the grounds that “there is likely to be relatively little school non-response bias” (p248). The vagueness of the descriptions of the criteria applied in each instance is clearly problematic, particularly as the means of assessing bias in each instance were not commensurate and, on a strict reading of the inclusion criteria, both should have been excluded from the tables for failing to obtain at least a 65% pre-replacement response rate. The UK study was able to gain strong leverage on pupil level nonresponse bias by merging in contemporaneous achievement data at the

⁵ It is not clear from the PISA technical reports whether the pupil level response rate criterion of 80% relates to the pre or the post replacement response rate.

individual level, while the US had no such pupil level data that was highly correlated with achievement. Thus, it is entirely possible that the failure to find evidence of nonresponse bias in the US data was a function of the weakness of the nonresponse study, rather than the absence of nonresponse bias. If OECD wishes to achieve a determination which is 'transparent, based on evidence and defensible' in future rounds of PISA, it is essential that the criteria used to make a determination with regard to acceptable levels of nonresponse bias are set out clearly, with accompanying rationale, in an openly available publication prior to the process of adjudication.

6.2 The Use of Replacement Sampling

Replacement, or 'substitution', is highly prevalent as a sampling strategy in many countries throughout the world. However, its 'acceptability' in the survey methodological literature depends on both the exact form in which it is implemented and the way that response rates are calculated post-replacement. Generally, however, substitution is regarded with a strong degree of scepticism in the survey methodological literature and is rarely advocated in survey textbooks (Vehovar 1999). Lynn (2004) notes up to 12 different varieties of replacement, defined as a function of the stage at which replacement occurs, whether the interviewer or 'office' makes the selection, and the means of selecting the replacement units. In terms of Lynn's classification, the PISA design would be classified as type 8 in which (1) the office makes the decision over whether to issue a replacement unit, (2) the office selects the replacement unit, and (3) the replacement unit is selected via stratified random sampling. There is nothing inherently wrong with this procedure, it is in fact commensurate with Kish's 'supplement samples' procedure, in which a 'reserve' sample is drawn at the first stage of sampling and issued only in the event of failing to meet a pre-specified sample size from the original or 'first' sample. The problem most often identified

with type 8 substitution is when the substituted units⁶ are not included in the denominator of the response rate calculation, rendering this statistic misleading as an indicator of the cooperation rate and, thereby, survey quality (Lynn 2004). As Vehovar (1999) notes, perhaps the primary danger of substitution is the potential for it to create “the illusion that nonresponse has been removed” (1999, p.348). Additionally, where interviewers are aware of the possibility of substituting nonresponding units, they may make less effort to persuade initially sampled units to cooperate. This has the potential to introduce bias into survey estimates as a direct result of the substitution (Chapman 1983). The most authoritative guide to response rate calculation, The American Association of Public Opinion Research’s ‘Standard Definitions for Final Dispositions of Case Codes and Outcome Rates for Surveys’ states, with regard to substitution:

“All replaced cases must be accounted for in the final disposition codes. For example, if a household refuses, no one is reached at an initial substitute household, and an interview is completed at a second substitute household, then the total number of cases would increase by two and the three cases would be listed as one refusal, one no one at residence, and one interview”.

(AAPOR 2004).

Although the PISA response rate tables do present pre and post-replacement response rates, the latter are calculated by excluding nonresponding schools that were successfully replaced by a substitute, from the denominator of the equation.

⁶ A substituted unit constitutes a unit which declined to participate and was successfully replaced by a first or second substitute. It does not include replacements for participating units in the initial sample.

We believe this contravenes best-practice and presents a misleading indicator of the overall cooperation rate in the survey.

The rationale behind replacement sampling **as a means of bias reduction** in the PISA survey is that nonresponding schools may be considered equivalent to the substitute schools on the matching variables. If the stratification variables are correlated with the survey variable of interest, then replacing nonresponding schools with substitutes is more or less equivalent to post-stratification via the matching variables (Elliot 1993). For example, in the 2003 UK PISA survey, sampled schools were matched to replacements on school size, school type, exam result category, gender mix (for independent schools) and LEA. This is because matching schools were identified by selecting the schools immediately before and after the selected school in the stratified sampling frame. For the use of replacement schools to remove rather than simply reduce nonresponse bias, response propensity and pupil achievement must be uncorrelated, conditional on the matching variables. It is easy, however, to conjecture about additional variables which might violate this assumption. For instance, 'head teacher's subjective assessment of pupil competencies' is a school characteristic which is not used in the matching procedure, yet is likely to be correlated with both participation rate and test performance. Therefore, a survey with a 100% post-replacement response rate could still produce biased estimates of pupil achievement, a possibility which highlights the misleading nature of the post-replacement response rate formula employed by PISA.

It would be relatively straightforward to investigate the effectiveness of replacement schools in reducing bias in estimates of pupil achievement. Because bias can be reliably estimated via weighting adjustments using GCSE scores of responding and nonresponding pupils (Micklewright and Schnepf, 2005), the effect

of replacement schools on the degree of bias could be estimated by differencing the bias at first and second phase. A more serious concern about replacement sampling than its deficiencies as a bias reduction strategy, however, is its potential for introducing bias into the sample. Where fieldwork procedures are exactly equivalent between the initial and replacement units, substitution will have no effect on the bias. However, such equivalence is rarely, if ever, observed in practice. Often, there are differences in, *inter alia*, the length of the fieldwork period, interviewing staff employed and deployment of refusal conversion, which make the fieldwork conditions non-equivalent. Such non-equivalence is often manifested as differences in the contact and cooperation rates between the initial and replacement samples, as, indeed was the case for PISA 2003. It is easy to see how such differences in fieldwork characteristics can introduce bias to survey estimates. For instance, if the fieldwork period for the replacement sample is shorter than was the case for the initial sample, then it is likely that fewer contacts will be made and less overall effort expended by the survey organisation in trying to recruit replacement schools relative to the initial sample. Replacement schools, once contacted, may also be less willing to participate if they are offered a smaller window within which to organise and conduct the tests. With PISA there is the additional complication that replacement schools may be asked to test pupils at what might be considered a more 'difficult' time in the school year. All of these factors are likely to result in the replacement sample containing a larger proportion of 'willing' schools and a lower proportion of 'unwilling' schools relative to the initial sample. If willingness to participate in PISA is correlated with pupil achievement, this would mean that the practice of replacement sampling would be a direct cause of bias in survey estimates.

The use of replacement sampling has been the source of an ongoing dispute between Prais (2003; 2004; 2005) and the OECD (Adams, 2003), with the former arguing it contravenes consensually accepted principles of statistical sampling, being more akin to quota sampling, and the latter justifying its use as both acceptable and standard practice in the field. While noting that these exchanges have tended to generate rather more heat than light, we agree with Prais' primary grievance that, in calculating response rates after the use of a replacement sample, the replacement schools should be included in both the numerator and denominator of the equation (that is approach (ii) in Prais (2003)). We find the argument that this approach "does not reflect the principles underlying the use of replacement samples, which attempt to reduce non-response bias" (Adams, 2003, p.385) to be both tautological and founded on the untested and implausible assumption that "replacement schools are perfect matches for non-responding schools" (ibid).⁷ Although replacement might serve to reduce bias in some cases, in others it may very well introduce bias where none existed in the initial sample. Replacement, furthermore, is certainly not a cost effective strategy because, as noted above, the same effect can be attained through the application of post-stratification weighting adjustments or imputation via the matching variables (Elliot 1993; Lynn 2004). And, while replacement might be preferred to post-stratification on the grounds that it is not subject to potential precision losses as a function of

⁷ The question of whether replacement sampling is akin to quota sampling is tangential to the central issue of bias reduction and the calculation of response rates. That said, however, both authors are, to a degree, correct. Quota sampling effectively substitutes replacement respondents for those declining to participate, assuming zero correlation between response propensity and survey variables, conditional on the quota controls. However, as Adams notes, quota samples are not drawn randomly from a sampling frame, meaning the probability of selection is unknown.

variance in the weights, this must be pitted against the fact that replacement runs the risk of failing to fully adjust for differential nonresponse. That is to say, certain strata may still be under-represented after replacement due to nonresponse amongst the substitute units. Any such residual would need to be compensated by post-stratification anyway, as is the case in PISA. A final reason for preferring post-stratification over replacement is that the former can sometimes make use of additional variables that were not used in the initial stratification. For example, GCSE scores are much more strongly correlated with PISA achievement than any of the variables used for stratification in England but were not available at the time the sample was drawn. This means that a weighting adjustment using the GCSE scores would be considerably more effective as a means of bias reduction than replacement via the stratification variables alone⁸.

In sum, then, it is our view that replacement, as implemented in the PISA survey, is a useful means of ensuring a sample size target but is of dubious utility and cost-effectiveness as a means of reducing bias. Because differences in fieldwork conditions between the initial and replacement sample may serve to introduce bias, a safer means of ensuring a pre-determined sample size would be to use Kish's supplement sample procedure (Kish 1965) in conjunction with post-stratification to correct for differential nonresponse. Where replacement sampling is used, the response rate formula should include substituted units in the denominator in order to be an accurate indicator of survey accuracy.

⁸ Although, of course, any additional weighting variables would need to be available to allow timely publication of the comparative tables.

6.3 Response Rate Thresholds

The limitations of replacement sampling are, furthermore, recognised in the PISA sampling documentation, “The use of replacement schools does not guarantee that potential biases have been reduced”⁹ and NPMs are encouraged to attempt as high a response rate as possible at phase one to limit its use. In addition to verbal discouragement of its use, the PISA sampling strategy explicitly penalises the use of replacement schools by implementing a sliding threshold for acceptable post-replacement response rate. The threshold weighted school response rate increases linearly as a function of the number of replacement schools used in the achieved sample. The formula for determining the threshold ‘after replacement’ response rate is given as¹⁰:

$$y=(255-x)/2$$

, where y is the threshold ‘after replacement’ response rate and x is the first phase (pre-replacement) response rate. Table 6.1 shows a range of pre and post replacement response rates using this formula. We have been unable to identify a published account of the exact rationale underlying this formula. However, it effectively applies a penalty such that, for each percent below the first phase threshold a country’s response rate falls, the post-replacement threshold is increased by half a percent.

⁹ School Sampling Preparation Manual, PISA 2003 Main Study, p22.

¹⁰ This formula can be found in the document ‘Technical Standards for the Implementation of PISA 2003’, p.5.

Table 6.1 Threshold Response Rates after Replacement

first phase	Threshold after replacement
80.0%	87.5%
75.0%	90.0%
70.0%	92.5%
65.0%	95.0%
60.0%	97.5%
55.0%	100.0%

Again, we would note that there is no strong reason to assume that bias operates in this neat linear manner; it is perfectly possible that the use of replacement schools completely removes any first phase nonresponse bias, with the use of the sliding threshold effectively serving to disqualify unbiased data. No distinction is made in this formula between whether a replacement school is a first or second substitute, although following the logic of increasing the post-replacement threshold, the same penalty should presumably apply when a second substitute replaces a first substitute.

Another important implication of this sliding 'post-replacement' threshold is that we can identify the likelihood of making the 'after replacement' threshold on the basis of the first phase rate. This we do by assuming that the second phase response rate will be the same, or at least no better, than the first phase rate (this was, indeed, true of the 2003 PISA fieldwork in the UK). This is because we assume the fieldwork agency will have left no room for improvement in their efforts to maximise first phase response. In practice, the shorter lead time and other differences in fieldwork conditions for the replacement sample means that cooperation is likely to be lower at this stage.

The ability to identify the likelihood of meeting the post-replacement threshold on the basis of the first phase response rate is beneficial in at least two respects. First, it might serve to present a less ‘daunting’ hurdle to be cleared at first phase; believing a target response rate is achievable at the outset should boost organisational motivation and morale in the data collection agency. Second, being able to identify that meeting the post-replacement benchmark is extremely unlikely on the basis of the first phase response rate could prevent the needless expenditure of public funds on further data collection and the concomitant waste of scarce school and pupil time¹¹. Table 6.2 shows, for a hypothetical country issuing 400 schools at the first phase, a range of first phase response rates along with the required response rate in the replacement schools sample. We begin by considering only the first substitute schools because, where fieldwork is conducted in the spring term (as per PISA 2000 and 2003), it is unlikely that sufficient time would be available for a third phase of interviewing.

If we look at the first row of Table 6.2 we see that at the first phase, a response rate of 80% was achieved, 5% short of the first phase benchmark. As 400 schools were issued, this equates to an achieved sample of 320 schools. Because of the 5% shortfall, the post-replacement response rate benchmark becomes 87.5%

¹¹ There is, of course, the possibility that the data might be included in the published tables without reaching the post-replacement threshold, on the basis of a bias assessment presented to the PISA sampling referee. This might strengthen the case for proceeding with the replacement sample, even if the analysis shows that it is unlikely that the post-replacement benchmark will be met. However, as we have noted in section 6.2, replacement sampling may, in fact, serve to exacerbate rather than reduce any existing nonresponse bias, so careful consideration should be given to this before proceeding with a second phase on the basis of this rationale. .

(=255-80/2). A replacement school is issued for each nonresponding school at the first phase, 80 in total.

Table 6.2 Second Phase Response Rates as a Function of First Phase Rate

1 st Phase Response	1 st Phase Issued	1 st Phase Achieved	Threshold After replacement	Replacement schools Issued	Number required	Rate	Total Achieved
80.0%	400	320	87.5%	80	30	37.5%	350
75.0%	400	300	90.0%	100	60	60.0%	360
74.0%	400	296	90.5%	104	66	63.5%	362
73.0%	400	292	91.0%	108	72	66.7%	364
72.0%	400	288	91.5%	112	78	69.6%	366
71.0%	400	284	92.0%	116	84	72.4%	368
70.0%	400	280	92.5%	120	90	75.0%	370
69.0%	400	276	93.0%	124	96	77.4%	372
68.0%	400	272	93.5%	128	102	79.7%	374
67.0%	400	268	94.0%	132	108	81.8%	376
66.0%	400	264	94.5%	136	114	83.8%	378
65.0%	400	260	95.0%	140	120	85.7%	380
64.0%	400	256	95.5%	144	126	87.5%	382
63.0%	400	252	96.0%	148	132	89.2%	384
62.0%	400	248	96.5%	152	138	90.8%	386
61.0%	400	244	97.0%	156	144	92.3%	388
60.0%	400	240	97.5%	160	150	93.8%	390

To meet the post-replacement benchmark response rate of 87.5%, a total of 350 schools will need to be achieved from across the two phases. As 320 schools were achieved at the first phase, this means that a total of 30 schools need to be achieved from the 80 replacement schools issued at phase 2, a response rate of 37.5% for phase 2. Given our assumption of equal response rates across phases, this should be eminently achievable. If we continue down the rows of Table 6.2 we find that only when the first phase response rate falls below 72% does the

second phase rate increase above the level achieved at phase 1. At this stage, the post-replacement benchmark will not be reached without somehow violating the assumption of equal response rates across the two phases.

Table 6.2 makes no allowance for the use of second substitute schools and assumes equal response rates at each fieldwork phase. If we relax the latter assumption and introduce second substitute schools as a third phase of fieldwork, the PISA response rate targets become less daunting still. Table 6.3 shows post-replacement response rates for first and second substitute samples under a range of first phase response rates, assuming a 5% drop in response rate at each successive fieldwork stage. So, for instance, where a response rate of 68% is achieved at phase 1, we assume a response rate of 63% for first replacement schools and 58% for second replacement schools. This results in a post-replacement response rate of 95% at the third phase of fieldwork, which exceeds the PISA post-replacement threshold of 93.5%. Under these assumptions, then, the minimum response rate required at phase one becomes a considerably more achievable 68%.

Table 6.3 Target Response Rates using second substitutes

First phase PISA	1st replacements		2nd replacements		
	Response Rate	After repl. target	response rate	post-replacement rate	response rate
80.0%	87.5%	75.0%	95.0%	-	-
75.0%	90.0%	70.0%	92.5%	-	-
74.0%	90.5%	69.0%	91.9%	-	-
73.0%	91.0%	68.0%	91.4%	-	-
72.0%	91.5%	67.0%	90.8%	62.0%	96.5%
71.0%	92.0%	66.0%	90.1%	61.0%	96.2%
70.0%	92.5%	65.0%	89.5%	60.0%	95.8%
69.0%	93.0%	64.0%	88.8%	59.0%	95.4%
68.0%	93.5%	63.0%	88.2%	58.0%	95.0%
67.0%	94.0%	62.0%	87.5%	57.0%	94.6%
66.0%	94.5%	61.0%	86.7%	56.0%	94.2%
65.0%	95.0%	60.0%	86.0%	55.0%	93.7%
64.0%	95.5%	59.0%	85.2%	54.0%	93.2%
63.0%	96.0%	58.0%	84.5%	53.0%	92.7%
62.0%	96.5%	57.0%	83.7%	52.0%	92.2%
61.0%	97.0%	56.0%	82.8%	51.0%	91.6%
60.0%	97.5%	55.0%	82.0%	50.0%	91.0%
55.0%	100.0%	50.0%	77.5%	45.0%	87.6%

7 QUALITATIVE INTERVIEWS WITH NONRESPONDING SCHOOLS TO PISA 2003

7.1 Methodology

A list of 80 schools who had not participated in PISA 2003 was provided by the DfES. Schools were approached on a rolling basis until twenty five interviews had been completed. Interviews were conducted via the telephone during May and June 2005. Interviews were transcribed and analysed using the Framework method (Ritchie and Lewis 2003). The advance letter sent to schools and the topic guide used to conduct the interviews can be found in Appendix X.

7.2 Summary of Key Findings

- Head teachers are generally positive about and supportive of school based survey research. However, this general view is undermined by the sheer volume of requests they receive from a wide array of sources, including government, universities, market research organisations, and professional teaching associations, to name just the primary sources.
- These requests come in addition to statutory requirements to provide statistical information to a range of other bodies such as Local Education Authorities and DfES. Heads frequently expressed the belief that requests are repeatedly made for the same basic information and that much of this information should already be available from other sources. Thus, survey requests are seen as unplanned, uncoordinated and frequently redundant with one another.
- Survey requests must compete with higher priorities, not least of which, of course, are activities focusing on teaching and learning. Many heads stated that they are judged purely and simply on the academic standards in their school and that they must, therefore, think very hard before committing to anything which detracts from this focus. Taking pupils out of

lessons to take the PISA tests, or participate in other research, could be at the detriment of their performance in the tests on which they and the school are judged.

- Given the high level of pupil testing in schools these days, there is a general reluctance to submit pupils to further academic audit that is not a statutory requirement. Heads are, in short, faced with a difficult job of prioritising amongst the vast number of different survey requests they receive. Compounding this is the reduced availability of staff to coordinate survey requirements as a result of increasing teacher workloads.
- In making judgements about which pieces of research to participate in, the over-riding criterion expressed in the interviews was the degree of benefit and relevance to the particular school in question.
- While Heads could see the more general benefit of international comparative surveys such as PISA, they found it hard to see how participating would aid them in their key role of raising standards in their own schools.
- There was general support for the provision of a monetary payment for participation in a burdensome survey such as PISA, although this should not be seen as a general panacea; some schools stated they would not participate even if the incentive were raised to a five figure sum.
- Provision of feedback was frequently mentioned as something that might encourage participation. However, schools would not be so interested in receiving information about where the UK stands in an international league table, the kind of feedback desired would enable Heads to see how their own pupils were comparing nationally and internationally in key subject areas.

- The time of year at which PISA is implemented was not seen as something that would have a significant impact on the likelihood of participation in school surveys. Many Heads expressed the view that there was never a good time to implement such a burdensome survey. Although some felt that spring was a particularly bad time due to impending GCSEs, others noted that switching to testing in the autumn would simply bring a different set of problems.

The following sections of the report present our analysis of the interview data in more detail with supporting quotations. First we focus on more general issues regarding survey research in schools. We then move on to look more specifically at awareness and perceptions of the PISA survey.

7.3 General feelings about School-based research

In general head teachers recognised the importance of school research and were positive about the potential to learn from the findings. Research that could provide robust data to inform the way schools are run and how they compare nationally and internationally was welcomed.

“We need research, any research in a learning organisation you need research to inform you of the strategic planning and evaluation” (Int 18)

“I believe it’s got great value for, well for education and school leaders and for teachers themselves in terms of developing and reflecting upon practice really.” (Int 19)

“I think it’s about informing what comes next in education.” (Int 08)

Involvement in research conducted by DfES was deemed ‘essential’ by some head teachers believing it would lead to better informed decision making.

“The information that we can feed along to the DFES or share with them the better it is for us really... So my view is that I think it’s essential that we get involved in them.” (Int 19)

Although head teachers perceive clear system benefits from participating in school-based research they have to say ‘no’ to the majority of requests.

“By and large if I’ve got the capacity to respond we do as a school. It’s not that we are against surveys at all. I think they have got a place, but it’s about capacity...Surveys go to the bottom of the pile of your things to do.” (Int 08)

“We try and do as much as we possibly can. There are times when I’ve

said no because of time of year, when people are exhausted and things like that, or surveys that I think are badly worded or organised.” (Int 19)

So, Head teachers are not viscerally unwilling or averse to participating in school surveys but are subject to a number of competing priorities and time constraints which mean external research studies, by necessity, come low down on the agenda.

Heads report receiving a huge number of requests to participate in research, with the sheer quantity proving an unwanted administrative burden. They express concern that schools are being ‘over-surveyed’ and a strong belief that the number of research requests has increased in recent years.

“I ought to get my secretary to keep a log of it. I mean at times it is almost a daily occurrence” (Int 04)

“I would say virtually every other week we are doing a survey of some sort or another” (Int 19)

“I would say that the last 3 or 4 years has got worse. Before then I hardly ever got asked anything. So I suppose it is government trying to be more inclusive but it seems to have gone from nothing to manic really in a very short space of time. A pain in the neck” (Int 09)

“It has increased in the last few years” (Int 03)

Schools report receiving survey requests from a wide variety of sources, ranging from individual university students, LEAS, DfES through international surveys conducted for the OECD. They find it hard to prioritise one survey over another, as all claim to be of the greatest importance.

“Well, from universities, from Institute of Education, from research students individually, I mean I must get dozens of them every year...certainly marketing organisations” (Int 9)

“One of the problems is those people who send a survey to you take no account of all the other surveys that we are being sent as well. Everyone just assumes that theirs is the only issue” (Int 21)

As well as surveys and other research requests received by schools, Heads are also regularly required to provide administrative data for their LEA and the DfES. Heads often did not differentiate between these two categories, seeing them as part of the same undifferentiated mass.

“Remember that we’re also asked for information, not surveys but information from a huge plethora of people now as well. We’re always providing statistics for the LEA or the DfES or checking this or checking that, lists of everything, league tables of this, league tables of that.” (Int 09)

However, Head teachers were strongly of the view that they would not respond to surveys from political parties or participate in commercial research for private companies, especially if the research would be used to market or sell a product. Lower priority was also generally accorded to research conducted by university students.

Given the sheer weight of research requests, some heads felt they had begun to say no automatically, even to some worthy and relevant requests.

"We get so much of it that I don't want to do it any more you know, it's just gone beyond a joke really" (Int11)

"In that situation you tend to build up almost a blanket no and the danger there is you are saying no to some which are very, very important indeed" (Int16)

There were several factors which head teachers saw as influencing their decision on whether or not to participate in school based surveys. Of over-riding importance was the degree of relevance and direct benefit to their own school. Other important factors were the logistics involved in organising participation, teacher workloads, availability of requested information elsewhere and the timing of the research during the school year. These are considered in more detail below.

7.4 Relevance and Benefit

In terms of frequency, by far the most common reason for participation in survey research was given as relevance and benefit to the school; if the survey had clear potential benefit to their school and the students, Head teachers are often willing to participate. Relevance and benefit are seen as primarily relating to the teaching and learning activities and outcomes within their own school.

"The question I would ask is if there are outcomes from this survey are they going to inform me or others in such a way that we will be able to do a better or (more) efficient job for the students" (int 12)

"There's got to be some benefit for my students for me to want to do it" (Int 11)

I would actually say to the people in my school who are responsible for that area, 'if you took time away from the children's learning to do this, do you think you would get information from that that would be of direct benefit to you and to actually help the children?'" (Int 701)

Benefits could also be less directly focused on teaching and learning practices, such as the provision of monetary incentives and possible exemption from onerous administrative tasks. Several heads expressed the idea that there should be a *quid pro quo* for participating in research, such that they might not be required to participate in further research for the rest of the year, or could do the survey in lieu of statutory testing.

That would be an incentive, if there was an alternative, 'do these instead', well fine I wouldn't have a problem. 'Do these instead of SATS', I wouldn't have a problem but not 'do these and SATS'" (Int 701)

It was felt that if a questionnaire appealed to both staff and students they would be more likely to 'buy-in' to the research. Heads also noted that, for a survey request to be successful, the direct benefits of participation needed to be communicated upfront so head teachers and teaching staff could make an informed decision about whether to participate.

"It needs to be fairly clear that it's going to be of some relevance and some help in some way to the students or to the institution or to the teachers. It's got to be clear where you can benefit" (Int 10)

7.5 Detracting from Core Priorities

Even when heads recognise the benefits of research and regard a survey as relevant to their core activities, it always remains low down on the list of priorities. Head teachers and their teaching colleagues are committed to their core roles of teaching and students' educational needs always come first.

"I mean great, you know, a national survey, I'm all for them but at the end of the day I am paid to do the best for the children of (name removed) school". (Int 14).

"I can't find enough hours in the day to get them to do some of the things that are crucial...if it means sitting with a child or talking to a parent rather than filling in a questionnaire, I know what they'd choose every time." (Int 09)

"The only way you can do these surveys is actually to take children away from the job they should be doing" (Int 701).

When considering a piece of research like PISA a head teacher needs to balance the value and benefit to be gained from participating in the survey against competing priorities, including the operational running of the school as well student and staff attainment levels.

"The priorities would have been maintaining the smooth running of the school, acting on raising attainment and people trying to cope with their new responsibilities. So I don't think an external survey, however worthy, at that time would have fitted into the immediate and urgent priorities." (Int 29)

"There is no direct benefit to the learner and therefore why do it...there is an absolute benefit holistically but there isn't for the individual" (Int 16)

7.6 Monetary incentives

Most head teachers felt that monetary incentives were generally a good idea and particularly helpful to schools who are struggling financially. However, the view was also expressed that the financial incentives offered are often too small to make a real difference to the school and, as a result, are of little real incentive.

“Wow, that would be tremendous you know. Thirty pounds, forty pounds in terms of supply cover for that person. That would go down well and would be seen well by the member of staff concerned” (Int 05)

“I think if a payment was made it might make it easier because...it might help you to get some help or, you know, buy in a supply teacher” (Int 02)

“If there was significant resources or funding attached to it I might consider it more strongly. But the reality is the funding tends to be sort of tokenistic...so, no, it doesn't come into my thinking at all” (Int 01)

“£500 is money that I could very comfortably spend; it could go on something that would benefit either the staff or the girls...could possibly buy some new staging that we need in the Hall or something like that” (Int 23)

Others felt more strongly that no amount of money would persuade them to participate in school research, if participation would not be of benefit or relevance to their students or school.

“Paying us money and things like that ain't going to make any difference at all. It's not about money.” (Int 14)

“I mean, the financial incentive is always a good one, but it's not always practical and possible is it?” (Int 10)

7.7 Provision of feedback

Receiving feedback was felt to be an influencing factor in the decision to participate in school research.

“The best encouragement of all is that there will be something about those learners which you are going to get feedback on that could be helpful to you. That’s the thing that would motivate me.” (Int 16)

This was clearly linked to the notions of relevance and benefit; schools would not appreciate feedback from research which was not relevant to their core aims and activities.

“I don’t really want to know that 35% of students in England can do something because that still doesn’t tell me what my students can do. And fine, I don’t know where those students are that can do that” (Int 11)

Head teachers were disappointed when they were promised feedback which they subsequently never received. It is also felt to be important to clearly link the feedback received to the piece of research conducted; sometimes head teachers receive research reports but find it hard to remember which survey the findings relate to.

Head teachers express a desire to receive feedback focusing more specifically on the data emanating from their own school. Often, the feedback they receive is of a very general nature and consists of national level findings. While this data is interesting and often informative, head teachers would also like to know where the opinions of their own students and their own school sit within the broader trends. Such individual findings place the research in the context of the school and could feed directly into management of the individual school concerned.

Head teachers often feel that the findings they receive do not tell them anything they did not already know. Research that regurgitates what teachers and schools had told the researchers was felt to be of little use.

7.8 Repetition and Redundancy of Requests

Head teachers feel that many school surveys ask for the same information over and over again. Heads often feel that this information would likely be available from some other administrative source but that surveying schools is seen as the cheap and easy option. This results in a perception amongst Heads that schools are doing other peoples' work for them. This redundancy of requests for information is a cause of irritation for busy Head teachers.

"I kind of feel that we ought to give the information once and then people should be able to get the data from elsewhere." (Int 11)

"Stop asking for information that you, somebody in your department probably has somewhere" (Int 02)

Several Head teachers advocated a more co-ordinated approach be taken towards school based research in which all research requests emanate from a central source, which would manage the number of requests that schools receive each year.

"There is a lot of information now collected from schools and collected by LEAs about schools and about progress and it seems to me that all that information can be gathered from one central point rather than keep asking schools five or six, sometimes ten, twenty times to deliver the same information" (Int 14)

"There doesn't seem to be joined up thinking and maybe with this Government's plan for the single voice concept then maybe we should bring research into that as

well” (Int 18)

“Could it be rationalised so you are only hit every now and again?” (Int 9)

Some heads expressed the belief that participating in research simply increases the number of future request, as they are identified as being a ‘soft touch’.

“I say to my secretary, we must have a big arrow over our school saying we’ll do anything and the answer is ‘no, we won’t any more like we have in the past’. Its just too much” (Int 11)

Head teachers are less likely to take part in a major piece of research if they have already participated in a large survey during the same academic year.

“The other thing I would say is that having done a major survey this year I would not consider doing another one this academic year because I think you can just survey out...that’s why a lot of them are binned as well because you know it’s we did something similar last year or no I’ve just asked the girls to come out of lessons to do that survey and I’m not doing another one.” (Int 23)

7.9 Timing in the School Year

The timing of a survey within the school year was voiced as another factor which might influence a school’s decision on whether to participate in school based research. However, views vary about when in the school year was a good time to conduct research. Many argue that there is never a convenient time, as staff and pupils are always subject to competing time commitments.

“The year is frantic from beginning to end.” (Int12)

“There is no flat time now in a school year.” (Int 14)

Others believe that the best time to target students is very much dependent on which year group is completing the research. There are certain periods during the school year which are particularly difficult for students sitting their GCSEs or A-Levels. The advent of modular exams means that exam time is ‘ongoing’. Preparing for these exams also affects the teaching staff involved with these year groups.

Some head teachers pinpoint the Autumn and Winter terms as being the best time to target students and teachers. The beginning of term is thought to be a busy period but head teachers note that things settle down a few weeks into the term. The Summer terms is thought to be the worst time to conduct research on pupils, as most students and teachers are pre-occupied with exams.

“Well I have to say the summer term is an absolute no no because it’s pressure then. For the autumn term it’s probably the most convenient in terms of not having quite so many other things that we are having to meet deadlines for.” (Int23)

In contrast other head teachers believe that the winter months are the worst time in terms of student behaviour and that response rates would be lower as a result. These teachers felt that the summer terms were the best time to target students for the best responses.

"I mean obviously if it was May or June forget it because it's all exams, January is now out, as is the beginning of February because of AS's. Unfortunately with the kind of assessment regime that you have in schools now, there isn't really, I suppose July but then we all do trips. It's quite difficult to find a time when its ok to take kids out." (int 09)

One Head teacher commented that the best time to target Head teachers is at the beginning of the six week holiday in the summer. He speculated that at the start of this school holiday head teachers were more likely to have time to spare to sit down and complete a survey. This, of course, would not be suitable for surveys requiring pupil data.

7.10 Teacher Workload

Many heads report that teaching staff are always busy with competing pressures and that the sheer weight of teacher workloads means the human resource simply isn't there to do surveys any more.

"If we ask people to do other things on top, people haven't got the capacity. People are really stretched" (Int 701)

"I think that teachers as a profession obviously have tons of work to do and at certain times of year are really focussed on driving the attainment of their kids, so they get very worn out towards the end of term." (Int 19)

"I know many schools who have got inexperienced heads with inexperienced teams who are struggling...and actually to ask them to do something else is just sometimes the straw that breaks the camel's back." (Int 04)

"Teacher workload is the big problem. And assessment" (Int 28)

When teaching staff do have 'free' time this is usually taken up with additional work, including their own performance development work, training of new staff members or arranging extra curricular activities for their pupils, as well as the normal pressures of marking and preparing for lessons.

"To be quite honest with you like the majority of schools we are involved with so much work that when it comes to surveys we really don't get time to have a good look at it." (Int 20)

7.11 Awareness of PISA

Awareness of PISA varied from those who had never heard of the survey to those who had a good understanding of its purpose, main findings and exclusion from the published OECD Tables in the 2003 round. Some were totally unaware of PISA and claimed never to have heard of the survey and could not remember ever receiving the invitation to participate. This was, perhaps, surprising given that all schools in the sample had been contacted repeatedly with requests to participate in 2003. It is perhaps explicable though, given the sheer weight of survey participation requests received by schools and because some of the Heads interviewed for this study were not in post at the school at that time.

"I have seen it in the past" (Int 03)

"It's obviously important from the Government's point of view to have information that will actually help it make comparisons from nation to nation" (Int 701)

"Doesn't ring any bells at all" (Int 05)

"I'm afraid I can't remember the acronym, no, I'm afraid not" (Int 12)

Of those who had heard of PISA several noted that the survey was held in high regard by the government.

"Well I know that the Government rates it highly because it's a way of actually comparing us from society to society." (Int 04)

"We are probably the only country in Europe that are going to keep on getting our wrists slapped because we don't participate as a country and they can't get enough participation. I've seen that in the Press." (Int 23)

Even amongst those who had heard of PISA, few had an understanding of its primary aims, or the potential benefits of the UK's participation in it. It was described as an assessment of the key subjects (i.e. maths and literacy) across several different countries to allow for international comparison and benchmarking.

"My understanding is that it gives some measure as to the quality of learning and achievement in the basic core subjects across Europe. So it gives people involved in education a measure of how well they are achieving with their pupils in different countries." (Int 14)

7.12 Views on international comparative school surveys

When asked their views about international comparisons of student attainment more generally, opinions varied about the merits of this kind of exercise.

"So comparisons with the successes of say Finland and they're educational achievements, you know, there's some interesting data to be looked at." (Int 19)

"Come on, being blunt, what difference will it make where we are in a global league table?" (Int 08)

Others were concerned about the methodological and interpretational difficulties of making comparisons of attainment levels between students in England and students across Europe.

"If people don't know what's been measured or how it's been measured you know they can jump to all the wrong conclusions either accidentally or quite deliberately you know" (Int 04)

7.13 Views on PISA fieldwork requirements

The majority of Heads recalled the PISA requirements as being burdensome and disruptive.

"I seem to remember it was going to be asking you to take a selection of girls out of lessons and that requires us to go and make the selection, contact the girls, get them out of lessons, find a place for them to go to and I seem to remember somebody was going to come and actually do the survey." (Int 23)

"I mean I remember it meaning having to re-timetable kids. Have kids in different rooms at different times which actually all costs time and energy and money apart from the fact that it took away children from their direct learning." (Int 04)

"It just meant that someone's got to contact members of staff, contact children, you know its just time, it's just always about time." (Int 09)

"Oh, they were far too complicated" (Int 08)

Some Head teachers viewed the testing of pupils positively. A minority felt that it might be a good experience for students to get practice of sitting examinations.

"We thought it was a good opportunity for the children to have a rehearsal experience in doing a test." (Int 02)

The majority, however, felt that they did not want to have to put their pupils through 'yet another' assessment. There was also concern that, because PISA draws students from different form groups, the assessment would create a 'new' group that would require special supervision and accommodation.

"I mean we participate in random testing, so different kinds of tests before, so it wasn't that I was worried about the commitment it was actually I just don't want any more tests for the sake of tests." (Int 11)

"It's a disruption to teachers, it's a disruption to schemes of work. It's a disruption to rooms and basically to children not doing the work they should be doing." (Int 04)

"We are left with a group [of pupils] that doesn't exist outside of the survey and we have got to accommodate them. We have got to find them a supervisor, a room and the time." (int 26)

Head teachers appeared concerned about the timing of the survey and the age of the students selected to take part in the PISA survey. They pointed out that Year 11 students are quite possibly the worst year group to target, as they are involved in important exams as well as only being at school for two terms during the course of that year. Testing in the Spring was problematic as students were preparing for and anxious about their impending GCSEs.

"Year 11 is not the best year to survey because they are only in school for two terms because of exams." (Int 12)

"But by that time it was going to infringe on last minute preparation for exams or something like that. It was not at the right time of year." (Int 23)

"If I remember correctly, that was targeted at the most inappropriate time" (Int 03)

Others revealed that it was difficult to 'slot' PISA in to the school year due to existing commitments. Many head teachers felt that they did not receive sufficiently early warning of their selection in PISA to be able to slot it into the school's existing timetable of activities.

"We are under a lot of pressure from parents and the girls themselves and examination boards to meet deadlines, to get things done and we plan our programmes for teaching and the pastoral programmes for a long time in advance and to slot something in is actually quite hard. Along comes a survey and its not just a case of saying, we'll take those girls out of lessons. I can't do that." (Int 23)

Most could remember receiving letters explaining the survey and strongly encouraging co-operation. The series of letters relating to PISA were described as 'very polite pleading letters'. Others remember being offered escalating incentives to participate. Some wondered how much the amount might increase to if they kept holding out for more.

"I think I'm right in saying that the initial letter was a case of will you or won't you and I think we turned it down. We then had another communication later on, please, please, please, you know. Think of the greater good of England or something like that." (Int 23)

"I was saying to my secretary, 'I wonder how much they will offer before I finally say 'ok, right, do it'!'" (Int 04)

Some schools felt that the follow up letters and phone calls requesting participation were patronising and irritating. This was particularly so when the head teacher had made a carefully considered decision not to participate in PISA 2003.

"I wished they would get off my back." (Int 09)

"I had a discussion with an HMI who rang me up to say would I seriously consider it and I said, 'yes, and the answer is still no'." (Int 11)

"Yes I think I remember. I may have got a reminder about why I hadn't done it and why. That can be annoying you know, them saying you haven't done this, what

are the reasons.” (Int 18)

Some head teachers welcomed the idea of electronic completion of school surveys and believed this would encourage schools to participate. It was felt that surveys were quicker and easier to complete online. Others, in contrast, felt that online completion offered no obvious benefits and would be even more administratively difficult and would throw up new and perhaps more difficult technical and resource issues.

“Oh, I would always do that if I could. Personally, I don’t mind doing a survey electronically” (Int 23)

I don’t think so. In many ways, you know, the electronic thing can be even worse” (Int 04)

“Yes, but again it’s to do with resource if you’re talking about an area which would be used for learning not being available because kids are there doing a survey...” (Int 701)

8 SURVEY OF SCHOOL EXPERIENCES AND ATTITUDES TO SURVEY RESEARCH

8.1 Rationale and Sample Design

The qualitative interviews provide valuable insights into the ways that Head teachers evaluate requests for survey participation and the factors that are likely to influence their participation decision when requests are received. However, these interviews were based on a small and unrepresentative sample of schools which had been selected but not taken part in PISA 2003. It is possible, therefore, that the views elicited might not be representative of the total population of schools in England and Wales.

Additionally, therefore, we undertook a small quantitative survey of a random sample of English secondary schools, to gain a more robust and representative descriptive assessment of Head teachers' views. A simple random sample of 170 schools was drawn by DfES' sampling unit and passed to the BMRB telephone unit. Advance letters (see Appendix X) were sent to schools on DATE and interviewing began on DATE. Of the 170 schools approached, 87 provided complete interviews, 62 refused to participate and no contact was made with the remaining 21. This represents a response rate of 51%. The average length of completed interviews was nine minutes and twenty seconds. It should be noted that our achieved sample of 87 schools is too small to provide precise population estimates. The results of this study then should be seen as primarily descriptive and suggestive and any conclusions drawn as tentative and preliminary.

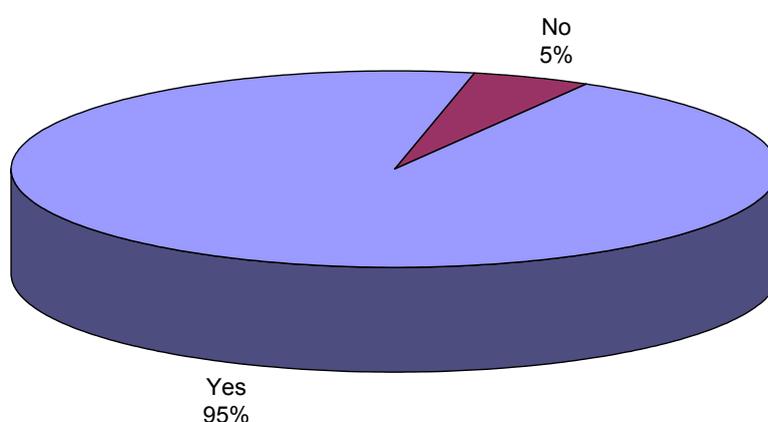
8.2 Results

The first questions in our survey examine the frequency with which schools receive requests to participate in research studies, the extent to which this has been changing over time and the experience and willingness of schools to

participate in survey research. Figure 8.1 shows that only a tiny minority of schools (5%) report having received no survey requests over the previous 12 months.

Figure 8.1

Apart from now, has your school received any requests to take part in surveys or other research during the past 12 months, that is since June 2004?

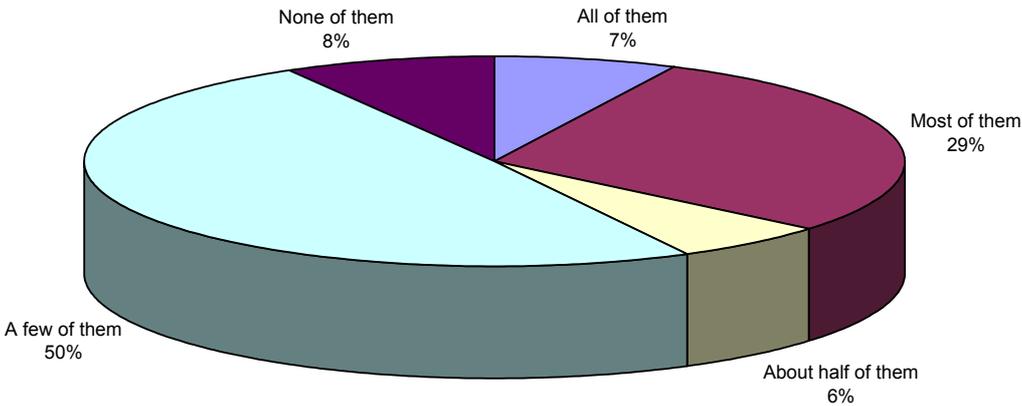


Those who reported receiving survey requests, were asked 'Approximately how many surveys or other research projects has your school been asked to take part in during the past 12 months, that is, since June 2004?'. The median response to this item was 10 requests over the previous 12 months. There was, however, a wide variation across schools in this estimate, with a sample standard deviation of 16. These are, of course, subjective assessments of frequency and may suffer from memory and various forms of projection bias. However, taking these estimates as they stand, the average secondary school in England and Wales is receiving somewhere in the region of one request for participation in some kind of research study every month.

Next, we asked how many of these requests schools had participated in. Figure 8.2 shows that only a small proportion (8%) had declined all requests received, with half reporting doing 'a few' and the remainder agreeing to the majority of requests. Interestingly, there is a significant correlation (.24, $p < 0.05$) between the number of requests received and the number of studies participated in; the more requests received, the lower the estimated proportion of times schools cooperated. We must, of course, be cautious in attributing a causal relationship here because it is perfectly possible that schools which are less likely to participate in research over-estimate the number of requests they receive. Nonetheless, this evidence provides some support for the idea noted in the qualitative interviews, that inundating schools with requests generates a general disinclination to participate in any single request.

Figure 8.2

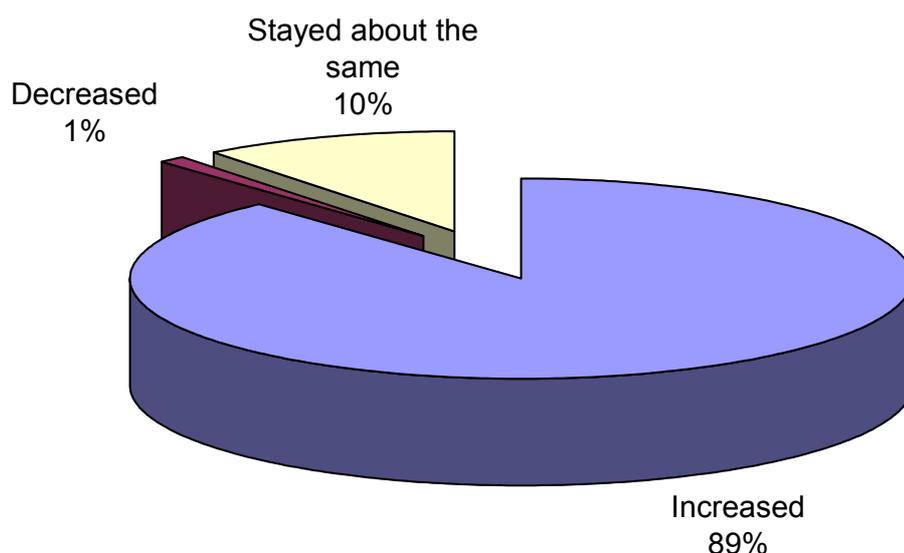
In how many of these surveys or other research projects did your school agree to take part?



We also asked schools whether the number of research participation requests had changed over the past five years (Figure 8.3). Overwhelmingly, schools reported that the number of requests had increased. Furthermore, of the 89% stating that

the number of research requests had increased, more than nine in ten (92%) stated that the number had 'increased a lot'.

Figure 8.3 Over the past five years would you say the number of requests to schools to take part in surveys or other research projects has...



In the last questions in the section of the survey pertaining to general requests for participation in research, schools were asked to provide verbatim responses to the following question, '*Over the past few years, schools have become less likely to take part in surveys or other research. What do you think is the single most important reason for this?*'. The six primary themes underlying responses to this question are summarised in Table 8.1 below. They are, unsurprisingly perhaps, strikingly similar to the results of the qualitative interviews presented in section 7 of this report. The most common reasons cited for schools being less willing to cooperate in research focus on the lack of time available, the perceived irrelevance of many research studies to schools core activities, the redundancy and repetition of requests, and the lack of useful feedback after participation. A

complete list of responses to this question for all respondents can be found in Appendix X.

Table 8.1 Most Common Reasons Given for Reduced School Cooperation

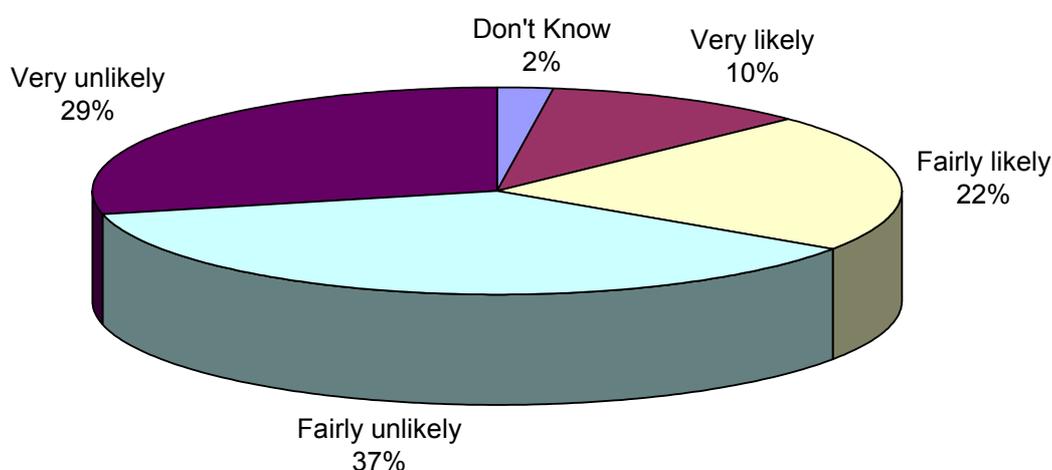
Lack of time/excess workload
Competing administrative requirements/Ofsted inspections
Lack of relevance/benefit for the school
Too many received, cannot prioritise one over another
Never receive feedback
Information requested already available elsewhere

The second half of the questionnaire asked respondents specifically about international surveys of pupil achievement such as PISA and TIMSS. Approximately half of schools reported not even being aware of the existence of such surveys (52%). Of those that were aware of these international pupil assessment surveys, approximately a quarter (28%) reported that their school had participated in such a survey, with the same proportion (24%) reporting that their school had declined to participate. This suggests that, where awareness there is of PISA and TIMSS, derives primarily from direct experience.

Next, schools were asked directly about the likelihood of their own school participating in this type of survey, if selected in the spring or summer term of 2006. Figure 4 shows that only a third of schools currently rate the chances of their school participating in this kind of survey, if selected, as 'likely', with the

remaining two thirds rating their probability of participation as either 'fairly' or 'very unlikely'.

Figure 4 'Suppose your school were asked by DfES to participate in an international survey of student achievement, which involved testing a random sample of 30 students from different classes, in years 10 and 11, for two hours in the Spring or Summer term next year. How likely is it that your school would agree to take part?'



Where schools reported that they were fairly or very unlikely to take part in such a survey, they were asked to provide, verbatim, their reasons for this assessment. Table 8.2 summarises the main reasons given by the 57 respondents asked this question. Again, there is a great deal of overlap with the results of the qualitative interviews and the verbatim responses given to the question concerning general reasons for declining willingness shown in Table 8.1. The only major difference from the reasons listed in Table 8.1 is the emphasis placed on this being a difficult time of year, especially for Year 11 pupils, as the majority will be preparing for impending GCSE examinations. A complete listing of the verbatim responses to this question can be found in Appendix ?.

Table 8.2 Most Common Reasons Given for being Unlikely to Participate in International Student Assessment Surveys

Bad time of year for this age group, competing pressure from GCSEs
Insufficient time/resources to organise it, teacher workloads
Students already 'over-tested'
Lack of direct benefit for school and/or pupils
Distraction from main task of teaching and learning

In addition to verbatim responses, Head teachers who rated the probability of their school participating as 'fairly' or 'very unlikely' were presented with a list of factors, drawn from the qualitative interviews, that seemed to be the strongest determinants of nonresponse. For each reason, they were asked to state whether the factor in question influenced their unwillingness to participate.

Table 8.3 shows that most of these reasons were overwhelmingly endorsed as relevant to the decision on whether to participate in this kind of survey. Only for the reasons relating to individual schools being 'targeted', survey results being irrelevant to schools, and surveys repeatedly asking the same questions did agreement drop below around 8 in 10 schools. Again, the clear message here is that surveys are seen as a distraction from the core activity of teaching, have little direct benefit to already over-tested pupils, are received too frequently and at inconvenient times in the school year.

Table 8.3 Reasons for Being Unlikely to Participate in PISA

Reason	% Agreeing
taking part would lead to too much extra work for staff at the school	83%
There would be no obvious benefit for the school	77%
It would disrupt the pupils' education	90%
Your school always seems to be targeted for this sort of thing	42%
There is too much testing of pupils in schools as it is	81%
You receive a lot of these requests, and you can't do all of them	86%
You do not find survey results helpful to your school or pupils	56%
Surveys always seem to ask you the same things	37%
Surveys often come at bad times in the school calendar	79%

Again drawing on the results of the qualitative interviews, all Head teachers were next presented with a list of actions and asked, for each one in turn, whether it would make them more likely to participate in a PISA style survey in 2006. Encouragingly, Table 8.4 shows that for all eight potential actions, Heads thought that they would be effective in making them more likely to participate in future rounds of this type of survey. The measures that would appear to offer the strongest leverage, at least from Head teachers' points of view, focus on the provision of resources to manage the survey, provision of useable feedback from the survey and omission from subsequent DfES survey requests for a specified period. Moving the testing session to a different time of year and providing a long

period of advance notice, although still seen as attractive, were rated as influential by a smaller proportion of Head teachers.

Table 8.4 Factors Increasing Probability of Participating in PISA 2006

Action	% Yes
our school were offered something in the regions of £500 to spend as it wished	64%
you were given 9 months notice of the survey	39%
DfES promised to send you no other non-statutory survey requests for a 12 month period	72%
additional temporary staff were provided to help your school prepare and administer the testing	68%
DfES provided funding for supply cover to free up time to prepare for the testing	60%
your school were provided with its survey results, showing how they compared with the national results	77%
test session were moved to a different time in the school year	51%
testing could be done on-screen using laptops provided for the purpose	64%

Finally, heads were asked to say what other things, not included in this list, might influence them to take part. Their responses focused primarily on improving the coordination, provision of useable feedback from the survey, reducing the overall burden of survey requests, provision of resources and exemption from statutory tests and/or future requests to participate in research. Some also mentioned the possible advantages of electronic test administration and the completion of tests in

pupils' homes rather than in school. A full listing of the verbatim responses to this question are provided in Appendix ?.

9 RECOMMENDATIONS

Drawing on the previous eight sections of the report, we make the following recommendations for improving response rates in school surveys in the UK. We begin with general recommendations for DfES' entire portfolio of research before moving on to actions more specifically germane to PISA and similar international surveys of pupil achievement. There is, necessarily, a degree of overlap between these two sets of recommendations.

DfES should set up an archive of meta-data on the research it conducts on its own account and through its external contractors. Such an archive would contain basic information on the aims of the study, the collectors of the data, target population, issued and achieved sample sizes and response rate. This would be of great benefit to DfES in keeping track of the quantity, quality and focus of the research it conducts and could be easily implemented as a standard requirement in all tendered contracts.

DfES should consider taking measures to limit the excessive number of requests to participate in research that UK schools are currently receiving. There is evidence that schools are beginning to decline to participate in research they themselves consider of great benefit, because they are too busy to invest the time in discriminating between the many requests they receive. It is our belief that schools would welcome the implementation of some sort of central clearing house to which those wishing to conduct school based research would need to apply as a first port of call. In addition to performing a rationalising and prioritising function, such a system might also usefully serve to reduce the large number of speculative and ill-thought out research studies currently conducted in UK schools.

DfES should examine the feasibility of rationalising sample designs across its portfolio of large, school based surveys. This might involve, for example, drawing an initial 'master' sample, from which sub-samples could be drawn for different surveys, with or without replacement, in a balanced and coordinated way. This would enable constraints to be placed on the maximum number of times individual schools can be selected within a school year. Such a system might also enable the implementation of *quid pro quo* agreements, whereby DfES could offer schools a guarantee of no further requests for participation within a fixed period, for cooperation in a particular survey or surveys.

For the 2006 round of PISA, a clear, adequately resourced, 'a priori' fieldwork strategy should be agreed at the outset between DfES and the data collection agency and. In terms of how fieldwork is implemented, we make the following recommendations:

The data collection agency which wins the tender to conduct fieldwork for PISA 2006 should be required to explicitly state what measures it has in place to ensure that turnover of key staff working on the survey is kept to a minimum. Additionally, a clear replacement/shadow management structure should be in place in the event of unavoidable staff departures.

DfES should seek to obtain permission from OECD to conduct PISA fieldwork during the autumn, rather than the spring term. Although schools are still extremely busy at this time of year, GCSEs will not quite be imminent at this time for the target age group. An additional advantage of conducting fieldwork at this time is that, should the field period need to be extended for second and third phases of replacement schools – as happened in 2003 – the extension would be

in a period during which most 15 year olds are still in school. After the Easter holidays many pupils have, *de facto*, left school.

The sample should be drawn as soon as is feasibly possible under the technical conditions set down by PISA. Contact should be made with schools to inform them that they have been selected to participate in PISA, as soon as the sample has been drawn.

The initial contact to schools should be in the form of a brief letter requesting a short personal meeting between the relevant 'gatekeeper', who will usually be the Head, and an experienced interviewer, whose role will be to 'sell' the survey and address any reservations expressed by the Head.

The initial contact letter should contain a substantial monetary 'incentive', paid to schools unconditionally in advance. It should be left to schools to decide how this money should be used. In addition, a monetary payment of at least an equivalent magnitude should be offered at this stage but conditional on the school's participation in the survey. Although the use of these monetary incentives will substantially increase the cost of PISA, it is our belief that the response rate benchmark is unlikely to be met without them. Thus, proceeding with PISA in the absence of a substantial monetary incentive may well prove a *less* cost-effective option for the UK taxpayer.

Our analysis of school surveys conducted by DfES in the UK over the past ten years revealed that only three percent achieved response rates of 85% or above (the PISA first phase benchmark) and these were considerably less burdensome than PISA. If, however, we make the plausible joint assumptions that the first phase rate will be unlikely to reach 85% and that the second phase rate will be

equal to the first phase rate, then the minimum first phase response rate needed to meet the post-replacement benchmark is 72%. Although still challenging, this can be considered eminently achievable with the deployment of sufficient resources. Assuming that both first and second substitute schools will be used and a drop in response rate of 5% at each phase of fieldwork renders a target response rate of only 68% for the initial sample. DfES and the fieldwork contractor should elaborate on this type of model under a range of different assumptions to determine the likelihood of meeting the PISA benchmarks under a range of different scenarios. Not only do we believe that adopting this perspective would aid in achieving benchmark response rate criteria, it might also prevent the potential waste of public funds on needless second and possibly third phase data collection.

DfES should take steps to raise the profile of PISA amongst head teachers in the UK. Our research indicates that only around a half of Heads have ever even heard of PISA and, of these, most are only familiar with it from having been included in the sample in the past. As part of this profile raising, DfES must be able to provide a clear exposition of its relevance and benefit to UK schools. At present, international achievement surveys are seen as 'worthy' but tangential to the core activities of individual schools.

Many sample designs for school surveys select schools with probability proportional to school size, as is the case with PISA. Given the frequency with which surveys are now being conducted on UK schools, there is a possibility that larger schools are being selected with great frequency, resulting in a general unwillingness to participate in survey (and other) research studies. DfES should

look more closely at the possibility that large schools are 'over-surveyed' as a result of their size.

The provision of feedback emerged from the qualitative and quantitative data as a significant factor in schools' decisions over whether to take part in research studies. However, they would not regard information regarding the international standing of UK schools as sufficiently relevant to their local contexts. DfES should, therefore, examine the feasibility of providing schools with feedback on how their students performed, relative to the other participating schools in the UK.

Earlier, we advocated the implementation of a rationalised sampling strategy across DfES' portfolio of large school surveys, with the possibility of placing upper limits on the number of requests received by schools within a particular academic year. We believe this might be of particular benefit to highly burdensome and irregular surveys such as PISA as Head teachers would consider the offer of 'no further surveys' as a prize worth participating for.

10 REFERENCES

American Association for Public Opinion Research. 2004. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 3rd edition. Lenexa, Kansas: AAPOR.

Atrostick, B. K., Bates, N., Burt, G., Silberstein, A., and Winters, F., (1999), "Nonresponse in U.S. Government Household Surveys: Consistent Measures and New Insights," Presented at *the International Conference on Survey Nonresponse, Portland, Oregon*.

Biemer, P. and Lyberg L. (2003) *Introduction to Survey Quality*. Hoboken, New Jersey: Wiley.

Bradburn, N. (1992), A Response to the Nonresponse Problem, 1992 AAPOR Presidential Address in *Public Opinion Quarterly*, 56(3), 391-397.

Chapman, D. (1983). The Impact of Substitutions on Survey Estimates. In *Incomplete Data in Sample Surveys, Vol. II, Theory and Bibliographies*, eds. W. Madow, I. Olkin, and D. Rubin, New York: National Academy of Sciences, Academic Press, 45-61.

Christianson, A., and Tortora, R. D., (1995), "Issues in Surveying Businesses: An International Survey," Chapter 14 in *Business Survey Methods*, edited by Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., New York: *John Wiley and Sons*.

Cox, B.G. and Chinnappa, B.N. (1995) Unique features of business surveys, pp. 1-17 in *Business Survey Methods*, edited by Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., New York: *John Wiley and Sons*.

de Leuw, E., and W. de Heer. 2001. Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In *Survey Nonresponse*, edited by R. Groves., D. Dillman, J. Eltinge and R. Little. New York: Wiley.

Elliot, D. (1993) The use of substitution in sampling, *Survey Methodology Bulletin*, #33, OPCS.

Fowler, F. (2002) *Survey Research Methods*, 3rd Edition. Thousand Oaks, California: Sage.

Goyder, J. (1987), *The Silent Minority*, Boulder, Colorado: Westview Press.

Groves, R.M.(1989), *Survey Errors and Survey Costs*, New York: John Wiley & Sons, Inc.

Groves, R., and M. Couper. 1998. *Nonresponse in Household Surveys*. New York: Wiley.

Groves, R, Floyd J. Fowler, M. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology*. New Jersey: Wiley.

James, J. M., and Bolstein, R., (1992), "Large Monetary Incentives and Their Effect on Mail Survey Response Rates," *Public Opinion Quarterly*, 56:442-453.

Jobber, D., (1986), "Improving Response Rates in Industrial Mail Surveys," *Industrial Marketing Management*, 15:183-195.

Keeter, Scott; Miller, Carolyn; Kohut, Andrew; Groves, Robert M., and Presser, Stanley, (2000) "Consequences of Reducing Nonresponse in a National Telephone Survey," *Public Opinion Quarterly*, 64, 125-148.

Lievesley, D. (1983), Reducing Unit Nonresponse in Interview Surveys, unpublished paper, London: SCPR.

Lipsey, M.W., and D.B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.

Lynn, P (2004) The Use of Substitution in Surveys, *The Survey Statistician*, 49, p14-16.

Martin, E. (2004) Presidential Address: Unfinished Business *Public Opinion Quarterly*, (68), 3, p439-450.

Merkle, D., and M. Edelman. 2002. Nonresponse in Exit Polls: A Comprehensive Analysis. In *Survey Nonresponse*, edited by R. Groves, D. A. Dillman, J. Eltinge and R. J. A. Little. New York: Wiley.

Micklewright, J. and Schnepf, A. (2005) PISA Bias Analysis, Draft Report.

Moore, D. L. and J. Tarnai. (2002) Evaluating Nonresponse error in Mail Surveys," In R. M. Groves, D. A. Dillman, J. L. Eltinge. and R. A. Little eds.. *Survey Nonresponse*, New York: Wiley.

NISS/ESSI Task Force on Participation Rates in International Assessments (2004) *Interim Recommendations*, National Institute of Statistical Sciences.

Ohlsson, E., (1995), "Coordination of Samples Using Permanent Random Numbers," Chapter 9 in *Business Survey Methods*, edited by Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., New York: *John Wiley and Sons*.

Paxson, M. C., Dillman, D. A., and Tarnai, J., (1995), "Improving Response to Business Mail Surveys," Chapter 17 in *Business Survey Methods*, edited by Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., New York: *John Wiley and Sons*.

Petroni, Rita , Richard Sigman, Diane Willimack, Steve Cohen, and Clyde Tucker. 2004. *Response Rates and Nonresponse in Establishment Surveys*. Paper read at Federal Economic Statistics Advisory Committee (FESAC) on . December 14, 2004.

PISA 2000 Technical Report, (2002) Organisation for Economic Cooperation and Development..

PISA 2003 Technical Report, (2005) Organisation for Economic Cooperation and Development..

PISA 2000 Technical Report, Organisation for Economic Cooperation and Development..

Platek, R. and Sarndal, C.E. (2001) Can a Statistician Deliver? *Journal of Official Statistics*, 17(1), 1-20.

Rasbash, John, Fiona Steele, W. Browne, and B. Prosser. 2004. *A user's guide to MLwiN version 2.0*. London: Institute of Education.

Ritchie, J. and Lewis, J. (2003) *Qualitative Research Practice; a Guide for Social Science Students and Researchers*, London, Sage.

Schleiffer, S. (1986), Trends in Attitudes Toward Survey Participation in Survey Research, *Public Opinion Quarterly*, 50, 17-26.

Scott, J.C. (1971), *Response Rate Trends 1955-1970 and Response by Season Size of Place of Residence and Length of Interview*, Technical Report 5a, Ann Arbor, Michigan: Survey Research Centre, Institute for Social Research, University of Michigan.

Srinath, K. P., and Carpenter, R. M., (1995), "Sampling Methods for Repeated Business Surveys," Chapter 10 in *Business Survey Methods*, edited by Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., New York: *John Wiley and Sons*.

Steeh, C.G. (1981), Trends in Nonresponse Rates 1952-1979, *Public Opinion Quarterly*, 45, 40-57.

Stoop, I. (2005) *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*, SCP, The Hague.

Vehovar, V. (1999) Field Substitution and Unit Nonresponse, *Journal of Official Statistics*, Vol. 15, No. 2, 1999, pp. 335±350

Voogt, R. (2004) "I'm not interested". Nonresponse bias, response bias and stimulus effects in election research. Amsterdam: University of Amsterdam.

Westat (2004) NISS/ESSI Task Force on Participation Rates in International Assessments – *Interim Recommendations* (Oct 11, 2004).

Wilcox, J. (1977), The Interaction of Refusal and Not-At-Home Sources of Nonresponse Bias, *Journal of Marketing Research*, 14, 592-597.

Willimack, D. Nichols, E. and Sudman, S. (2002) Understanding Unit and Item Nonresponse in Business Surveys. In *Survey Nonresponse*, Groves, R., Dillman, D. Eltinge, J. and Little, R. (eds), Wiley, New York.

11 APPENDIX

Copies of this publication can be obtained from:

DfES Publications
P.O. Box 5050
Sherwood Park
Annesley
Nottingham
NG15 0DJ

Tel: 0845 60 222 60
Fax: 0845 60 333 60
Minicom: 0845 60 555 60
Online: www.dfespublications.gov.uk

© BMRB International Limited 2006

Produced by the Department for Education and Skills

ISBN 1 84478 686 2
Ref No: RR721
www.dfes.go.uk/research