

Teacher surveys: The pros and cons of random probability surveys versus teacher panels

John Jerrim

Social Research Institute, University
College London, London, UK

Correspondence

John Jerrim, Social Research Institute,
University College London, 20 Bedford
Way, London WC1H 0AL, UK.

Email: jjerrim@ucl.ac.uk

Abstract

Two commonly used approaches to capturing information about teachers are random probability surveys and teacher panels. This paper reviews the strengths and limitations of these two approaches in the context of capturing information about the teacher workforce. A case study is then presented drawing upon recent teacher survey data collections in England. Although both designs should continue to play an important role in generating evidence about the teaching profession, random probability surveys of teachers should be used sparingly, and only when they will be properly resourced.

KEYWORDS

non-response, random probability samples, teacher panels, teachers

INTRODUCTION

Teachers are one of the most important inputs into young people's education and academic development (OECD, 2005). Replacing an 'average' teacher with a 'good' teacher is associated with a 0.2–0.3 standard deviation increase in standardised test scores each year (Hanushek, 2011). Yet, in many countries, there are ongoing challenges with teacher retention and recruitment. Given the importance of teachers—and the struggles to retain sufficient numbers in the profession—there continues to be great interest in this key occupational group. For instance, how satisfied are teachers in their job? How has this changed over time? How many expect to leave the teaching profession over the next 3 years, and what are their motivations for doing so? Understanding how the average teacher in a country or a region feels about such matters is critical to helping policy makers understand the concerns of the teaching profession.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Context and implications

Rationale for this study

Many teacher surveys are based upon random probability samples. Yet the major advantage of this approach over teacher panel data depends on obtaining a high response rate – which is rarely achieved.

Why the new findings matter

The paper walks readers through the pros and cons of teacher surveys based upon random probability samples as compared to using teacher panel surveys, drawing upon a case study of England.

Implications for governments and other organisations conducting teacher surveys

While random probability surveys of teachers should continue to be used, there should be great commitment to executing them properly. This means doing everything possible to obtain a high response rate. In many countries, including England, this has not recently been the case, with too many half-hearted attempts to gather data from a random sample of teachers that has ultimately ended in failure.

There are several potential sources of data that can aid in achieving this goal. Several studies have drawn upon general population surveys covering all adults in a country, identifying teachers via an occupational code¹ (e.g., Worth & Van den Brande, 2019). These resources have the advantage of allowing the characteristics of teachers to be compared to the broader adult population and to other occupational groups. For instance, Jerrim et al. (2020) draw upon such data to compare how the mental health of teachers compares to workers holding other professional jobs. However, the questions asked in general population surveys are not education specific, and thus cannot be used to understand key aspects of teaching that make it such a unique job (e.g., when it comes to workload, general population surveys are unable to distinguish between the time teachers spend on marking, lesson planning or liaising with parents). The sample of teachers in such resources also tends to be relatively small, particularly if one is interested in sub-groups (e.g., primary versus secondary teachers). Others have thus turned to administrative/register data about teachers as an alternative (e.g., the School Workforce Census in England). Although these resources are teacher specific and very large in size, they are not primarily designed for research purposes, and are therefore limited to objective information such as educational qualifications, sickness absence and demographic characteristics. Thus, critically, administrative records—such as general population surveys—are unable to garner the views and opinions of the teaching profession as a unique occupational group.

The third major type of resource—and the focus of this paper—are surveys that have been specifically designed to be completed by teachers, capturing important details about their lives and, in particular, their job. Such studies have the major advantage of being devoted to understanding the views of teachers and the vital role this group plays in their school and broader education system. They can thus capture teachers' views about key issues

they face in their job—such as marking and the behaviour of pupils—that make teaching distinct from many other occupations. Yet teacher-specific surveys also have limitations, particularly if their primary purpose is to provide an estimate of view of the typical/average teacher across the population (e.g., if one's goal is to estimate 'population parameters', such as average amount of time teachers in England spend marking each week). This is because teachers are a busy professional group who already have a lot of paperwork to complete and are increasingly being asked to participate in research. Thus, in turn, it makes it challenging to get teachers to complete surveys, particularly those that contain many questions.

There are, broadly speaking, two approaches used to survey teachers, including in England (the empirical setting of this paper). The first is where a random/probabilistic sample of all teachers within a certain geography (e.g., England) is drawn, who then typically complete a 30–45 min questionnaire. Recent examples of this design from England include the 2016 and 2019 Teacher Workload Surveys, the 2013 and 2018 Teaching and Learning International Study (TALIS), and the first wave of the Working Lives of Teachers and Leaders. The second set of resources are teacher panel studies. Rather than drawing a random probability sample of teachers from across the population of interest, these panels regularly ask questions to a set of willing participants (i.e., a convenience sample) over a period of time. The most prominent example in England—with over 8000 daily respondents—is Teacher-Tapp (<https://teachertapp.co.uk/>), although others include the NFER Teacher Voice (<https://www.nfer.ac.uk/publications-research/teacher-voice-omnibus-survey>) and subsamples of the YouGov panel (<https://yougov.co.uk/>).

Although both probability samples and teacher panels potentially offer great value in generating evidence about the teaching profession, there is sometimes confusion over their relative strengths and limitations. For instance, there is often an unfortunate tendency for some to equate a (drawn) probability sample to meaning 'better' or 'higher quality' data. Yet this may or may not be the case; critically, it depends upon how such surveys are executed and the research question(s) at hand. The primary goal of this paper is to help build a better understanding of these important issues in the hope of aiding government and other researchers make better (and more efficient) decisions regarding data collection from the teacher workforce. In particular, we hope to help readers understand when and under what conditions responses from a probabilistically drawn sample of teachers is likely to be the preferable approach and, on the flip side, when collecting data from a teacher panel is likely to be a better option.

To help achieve this goal, we present a case study of recent teacher surveys conducted in England. Several national surveys of teachers have been commissioned in this country over the last decade, many of which have drawn probabilistic samples. These all had the implicit intention of estimating population parameters (e.g., average working hours of teachers per week) and have indeed been used/interpreted in this way. However, as shall be discussed throughout this paper, there are serious doubts as to whether this has really been achieved. At the same time, teacher panels have emerged as a key source of data in England, and are increasingly being used (by government, teaching unions and academics) to inform education policy debates. Together, this makes England the ideal setting to consider the pros and cons of probabilistic teacher surveys as compared to teacher panels, and how evidence from both may be combined to generate the best possible insights into the teaching profession.

The paper now proceeds as follows. In the next section we focus on the strengths and limitations of probabilistic/random teacher surveys. An analogous discussion with respect to teacher panels then follows. Our case study of recent teacher surveys in England is then presented, and our conclusions are outlined in [the](#) final section.

RANDOM PROBABILITY SURVEYS OF TEACHERS

The goal of random probability surveys of teachers is to provide an estimate of what are known as ‘population parameters’—that is, to make generalised statements about the outcomes or attitudes of teachers from across the population (e.g., country) as a whole. The statistic of most interest is often about the average teacher in the country (e.g., average hours worked per week) but may also be the percentage of teachers in the population who fall above or below a specific threshold (e.g., the percentage of teachers who spend at least 5 h marking per week). To achieve this goal, a probabilistic sample is drawn. A cluster sampling approach is often used, with schools first randomly selected (often with probability proportional to size) and then teachers randomly selected to participate from within each school.² Occasionally, simple random sampling is used instead, where teachers are randomly selected to participate from a list of all eligible teachers within the population.³ The key assumption under either a cluster or simple random sampling approach is that the probability of inclusion in the sample is known a priori for each member of the population. The sample then selected to participate is then effectively a random subset of the broader population of interest.

Such a probabilistic sampling approach has three major benefits. First, as lists of teachers eligible to participate stem either from central government or school records, there is tight control over the population of interest and the potential set of respondents. In other words, we can be sure that the respondents will be teachers that meet the specific inclusion criteria set. Second, and perhaps most importantly, random sampling is the only way to have a high degree of confidence that one is obtaining unbiased estimates of population parameters—that is, that it is possible to extrapolate results from the sample to make generalised statements about the population of teachers as a whole. This is because—in expectation⁴—the random selection of schools and teachers will mean the sample will be very similar to the broader population in terms of both observable *and unobservable* characteristics. Finally, probabilistic sampling means it is also possible to quantify the uncertainty in the estimated population parameters from being based on a sample of teachers rather than a census. In other words, it is possible—and technically appropriate—to put a confidence interval around the results, providing an upper and lower bound of the true population parameter across all teachers in the population.

To be concrete, ‘observables’ in this context refers to any characteristics that we know the distribution of across the teacher population as a whole and also in our sample data—making it possible to verify the sample data are representative in terms of these characteristics (teacher gender, for instance). If it is not, then there are the statistical techniques that can correct for differences in terms of these observable factors (e.g., weighting). ‘Unobservables’ here refer to anything that we do not know the true distribution of across the population, and hence no independent information exists to verify the sample values against. The amount of time teachers spend on marking would, for instance, be an ‘unobservable’ characteristic in this context. If such unobservable factors influence whether a school/teacher responds to the survey and is also correlated with our survey measure(s) of interest, then this will lead to bias in our estimated population parameters. For instance, if teachers who spend a lot of time marking are less likely to complete the survey (e.g., because they are too busy with other paperwork) then this will lead the sample to underestimate the total average working hours of teaching in England. Critically, in contrast to observable differences between the sample and the population, there is little that can be done to detect whether such bias exists, let alone correct for it. In a truly random sample (with a 100% response rate) such selection into / out of completing the survey does not take place (e.g., teachers who spend a lot of time working are equally likely to be selected and respond to the survey as teachers who spend

very little time marking)—effectively ruling out the possibility that teachers with certain views or attributes are disproportionately over- or under-represented in the data.

A critical point to note is that it is the random selection of participants (teachers/schools) that ensures this benefit of random sampling is gained. It is this—and this alone—that allows one to effectively rule out the sample differing from the broader population of teachers in potentially important ways (that one cannot otherwise observe). Just as random assignment in experiments allows one to separate genuine treatment effects from potential observable and unobservable confounding factors, random sampling in surveys means the characteristics of the sample should (in expectation) be very similar to the population in all key observable and unobservable ways.

There is, however, one major threat to these theoretical gains from random sampling being realised—selective non-response. If teachers with certain characteristics do not respond to the survey, then the sample is no longer random. This, in turn, is likely to lead to biased estimates of population parameters—that is, one's estimates from the sample will no longer accurately capture the views of the average teacher across the population as a whole. For instance, if teachers who work particularly long hours are less likely to respond to teacher surveys (e.g., due to their lack of time), then estimates of average working hours from the sample will tend to *underestimate* the true average working hours of teachers across the population. Some have also argued that selective non-response means that standard methods used to quantify sampling error (standard errors, confidence intervals, significance tests) are no longer appropriate (Gorard, 2015).

What can be done to ensure the major theoretical benefits of random probability surveys of teachers are realised? The main strategy is to ensure that response rates to the survey are as high as possible. This is because any bias that gets induced into estimates of population parameters from non-random non-response is a function of:

- a. The 'selectivity' of the non-response. In other words, how different non-responding teachers are from responding teachers in terms of the attribute(s) the survey is attempting to measure.
- b. The amount of survey non-response. The percentage of initially sampled teachers that have not completed the questionnaire.

Very little can be done about point (a)—if we already knew the distribution of the attribute(s) of interest amongst responding and non-responding teachers, there would be little need for the survey in the first place. But, if point (b) can be limited—that is, a high response rate achieved—then any bias in one's estimates of the population parameters is likely to be minimal.

This then inevitably leads to the question—how high do response rates need to be? Unfortunately, there is no straightforward answer; response rates should be thought of as different shades of grey rather than being black and white. However, to offer some guidance, one can draw upon the criteria set by the OECD for participation in their Teaching and Learning International Study (TALIS). An important feature of this global teacher survey is that if the consortia conducting the study are not sufficiently convinced about the representativeness of the final sample, then results for that country are not included within the international report (or flagged as being problematic). As TALIS is a cluster random sample, the minimum response rate criteria in 2018 was set at 75% for schools and 75% for teachers. This meant that the overall minimum response rate requirement was around 50%⁵—that is, at least half of the initially randomly selected sample needed to take part.

Now, as Jerrim (2021) noted in the context of PISA, some countries do not meet the response rate criteria set in such international studies but are still included in the international reporting. Yet he also shows that—even when such criteria are met—the characteristics

of the achieved sample can still differ from known population values in important ways. Together, this suggests that a 50% overall response rate should be considered an absolute minimum that random probability surveys need to achieve. Otherwise, the major benefit of random sampling—being able to extrapolate findings to the broader population of interest—is at very high risk of being lost.

What can be done if a random probability survey of teachers has been conducted, but the response rate—despite best efforts—remains disappointingly low? Due to the high levels of non-response, representativeness of the sample in terms of unobservable characteristics can no longer be reasonably assumed. The best one can then do is to establish how representative the sample data are in terms of known observable characteristics of the teacher population.⁶ This can be done by comparing the sample of teachers one has managed to obtain data from to the broader population of teachers in terms of the characteristics that can be observed in both. For instance, just say we know the inspection rating of the school in which each sample participant works. One can then compare the distribution of inspection ratings amongst teachers in the sample to the known distribution across the population—that is, one can check if the sample obtained is at least ‘representative’ of the population in terms of this particular characteristic. The more school and teacher characteristics for which such comparisons can be made, the better. But it is particularly important—if possible—to compare the sample and population in terms of factors that are likely to be strongly associated with our survey questions of interest. Unfortunately, what is often known about the population of teachers is often limited to educational qualifications, job role, the school in which they work and demographic characteristics. Hence ‘balance’—that is, close correspondence between sample and population values in terms of a few selected characteristics—is likely to provide only limited reassurance of the representativeness of the sample obtained.

What if one performs such a comparison, but finds the sample of teachers differs from the population in non-trivial ways? In such situations, it is common to reweight the sample to try and improve its representativeness. Say, for instance, that one's sample is found to include only half as many teachers working in ‘Outstanding’ schools (based on its most recent inspection rating) as there are across the country. Weights could then be created so that teachers who work in Outstanding schools in the obtained sample are given greater emphasis when we produce our estimates of the population parameters (i.e., responses of teachers in Outstanding schools are effectively made to be worth double the responses provided by teachers working in non-Outstanding schools). This can, and often is, done to rebalance teacher samples in terms of observable background characteristics. Yet this approach can only correct for differences between the sample and population of teachers according to the limited number of factors that can be observed across both. Moreover, reweighting can also be done with teacher panel data as well. Hence, regardless of whether one reweights the data or not, the real potential gain from random sampling (to ensure the sample and population will also be similar in important *unobservable* ways) is unlikely to have been achieved due to the low response rate.⁷

There are also, of course, disadvantages associated with attempting to collect data from a random probability sample of teachers as well. [Table 1](#) provides an overarching summary, with each of the issues raised discussed in further detail as the paper progresses.

First, there is the issue of cost. Random sampling requires a sampling frame to be developed, the selection of participants to take place, the chosen schools and teachers to be approached to respond and then non-responders followed up. Each of these steps requires expertise and financial resource. Second, this sequential process takes time, with the questionnaire typically set well in advance of the survey taking place. This in turn means it is not possible for the questions to respond quickly to current events. Third, relatedly, random probability surveys of teachers do not usually occur very frequently—often just once per academic year (see the [Case study: Recent teacher data collections in England](#) section

TABLE 1 Summary of the advantages and disadvantages of teacher surveys based upon probability samples in comparison to teacher panels.

| | Random probability sample | | Teacher panel |
|---|---------------------------|------------------------|------------------------|
| | High response rate | Low response rate | |
| Representative of population in terms of unobservable characteristics | Likely. Unverifiable | Unlikely. Unverifiable | Unlikely. Unverifiable |
| Representative of population in terms of observable characteristics | Likely. Verifiable | Possible. Verifiable | Possible. Verifiable |
| Likely cost | High | Moderate | Low |
| Typical frequency | Annual | Annual | Termly to daily |
| Timeliness of data | Low | Low | High |
| Typical number of questions per year | 50 | 50 | Up to 1000 |
| Typical number of teachers | 2000–4000 | Variable | 1000–8000 |
| Provide longitudinal data | Very unlikely | Possibly | Yes |
| Measurement across academic year | No | No | Yes |
| Motivation/effort of respondents | Variable | Variable | High |
| Verifiable participants teachers | Yes | Yes | Indirectly |

Note: Each of the points covered in this table are discussed in further detail throughout the course of the paper—the intention of this table is to provide an overarching summary. Typical sample sizes, frequency number of questions and number of teachers are based upon recent teacher surveys conducted in England.

for further details with respect to the situation in England). This is likely to be a particularly important limitation if the policy environment is changing rapidly (e.g., during the COVID-19 pandemic) or if the attributes of interest—such as teacher workload—potentially vary over the course of an academic year. Fourth, to encourage high response rates, national probability surveys of teachers are often kept relatively short with only a limited number of questions asked (typically around 50 or less). Finally, securing high response rates may involve a trade-off with measurement error. For instance, although non-responding teachers may be followed up to encourage their participation in the study, there is little to ensure they put maximum effort into completing the survey if they do so. Thus, although high response rates may provide reassurance about the representativeness of the data, this could come at a cost of data quality suffering in other important (but harder to establish) ways.

TEACHER PANELS

In contrast to random probability surveys, teacher panels do not attempt to randomly recruit teachers from a broader population (e.g., country or region) of interest. Rather, they are formed of a convenience sample of willing recruits. These resources openly sign up teachers to the panel via various channels (e.g., social media, education events/conferences, word of mouth), and are free to respond to the questions posed as they please. Those who sign up to the panel typically answer questions on multiple occasions and are incentivised to do so (e.g., via prize draws, charity donations, vouchers). Although the frequency with which questions are asked varies across panels—see the [Case study: Recent teacher data collections in England](#) section for evidence on this matter in England—some such as TeacherTapp in England ask teachers questions every day.

Teacher panel surveys have several attractive features. One is that—unlike most random probability surveys—they provide longitudinal data about the same group of teachers over

time. They thus provide insight into how the views of a group of teachers are changing, including in response to key events. For instance, Jerrim et al. (2022) use data from a teacher panel to investigate how the work-related anxiety of a sample of teachers varied across 75 points during the COVID-19 pandemic. The longitudinal nature of the data collected also allows one to track the relationship between the views/opinions of teachers and their subsequent actions. For instance, teacher panel data can be used to explore how the job satisfaction of teachers at the start of the academic year predicts the probability that the teacher intends to leave their job (or the teaching profession altogether) at the end of the year.

A related benefit of teacher panels is that they can ask many more questions to teachers over time. For instance, the TeacherTapp panel are asked around 1000 questions each calendar year (three questions each day), compared to the norm of around 50 questions or less in a typical random probability teacher survey. This, in turn, means it is possible to build a much richer profile of survey participants. One can also track how different aspects of teacher's jobs vary across different days of the week and at different points in the academic year. For instance, how does the distribution of teacher workload across various tasks (e.g., marking, lesson planning, administration etc.) vary within and between the autumn, spring and summer terms? Teacher panels are well placed to provide evidence on such issues, whereas random probability surveys are not.

Another major attraction of teacher panels is that they are nimble and can quickly respond to important events. Say, for instance, the government announces a policy that will impact teachers and their schools. Respondents from teacher panels can provide almost immediate answers, sometimes within 24–48 h. An example is the recent teacher strikes in England regarding pay. When the government made an improved offer to teachers, data from teacher panels could quickly establish whether there was support for accepting the offer or not, and under what conditions (EDAPT, 2023). Clearly, the time-lag associated with conducting a high-quality random probability survey could not have achieved the same.

Finally, teacher panels are also likely to have motivated participants, as they have all willingly signed up to answer questions. As noted previously, although having disproportionately motivated respondents may impact upon data representativeness, it may improve data quality in other ways (e.g., respondents putting in greater effort and providing more considered responses). Many panels also provide opportunities for a wide array of individuals/organisations to pose questions to a large sample of teachers quickly and cost effectively, where otherwise this would not be possible.

On the other hand, the main disadvantage of teacher panels is the exact opposite of the main advantage of random probability surveys. Most importantly, the lack of random sampling precludes the possibility of ever being able to rule out there being unobservable differences between the sample and the broader population of teachers. Hence, just like a random probability sample of teachers with a low response rate, unobservable differences in the composition of the sample versus the population could always be offered as a potential (or at least partial) explanation for a particular result.⁸ Despite this, results from teacher panels are often used to provide (or interpreted as) approximate estimates of population parameters (i.e., are taken to represent what the 'average' teacher thinks). Likewise, as the sample is not a random selection from the population, often reported measures of uncertainty stemming from sampling variation (standard errors, confidence intervals, significance tests) are not technically appropriate (Gorard, 2015).⁹

The other key challenge with teacher panels is in ensuring respondents are indeed teachers. Whereas random probability surveys generate sampling frames based upon lists of teachers from government or school records, teacher panels rely upon self-reported status as a teacher. Although most teacher panels take steps to identify potential non-teachers via some of the questions asked, and checking consistency of their responses with external

information (e.g., the name and address of the school in which they work), control over who responds is often not as tight as in a random probability sample.

CASE STUDY: RECENT TEACHER DATA COLLECTIONS IN ENGLAND

The preceding section discussed how, in theory, both random probability samples of teachers and teacher panels have advantages and disadvantages. Yet, in practice, their relative merits depend upon how they are executed. In particular, the key advantages of random probability surveys hinge upon whether they provide more convincing evidence of being representative of the broader population of teachers, particularly in terms of unobservable characteristics (which, as discussed above, requires a high response rate). If not, then one has suffered the costs associated with conducting a random probability survey but without receiving the major gains. This section thus provides a case study from England to consider how random probability surveys of teachers compare to teacher panels in practice.

Random probability samples

[Table 2](#) provides a selection of national probability samples of teachers conducted in England over the last decade. These focus on large studies funded or conducted by government—mostly by England’s Department for Education. Together, these data collections have formed the main source of quantitative evidence used to inform recent teacher policy in England over recent years.

There is one common feature of these studies that immediately stands out. Although they include a large number of teachers—over 2000 in each—overall response rates are often very low. For instance, the 2016 and 2019 Teacher Workload Survey and the first wave of the Working Lives of Teachers and Leaders (conducted in 2022) achieved overall response rates of around 10%. In other words, for every 10 teachers that were part of the initial random sample, only one completed the questionnaire. For the teacher workload surveys, this was due to a combination of both school non-response (just one in four schools initially sampled agreed to take part) and non-response by teachers within the participating schools (even when a school agreed to take part, most teachers—around 60%—did not complete the questionnaire). With such poor response rates, the risk that the initial random selection mechanism has been broken is extremely high. Thus, despite random probability samples being drawn, these studies have in fact obtained data from what is essentially a convenience sample—that is, formed of a selected subset of willing participants. This, in turn, makes these studies unconvincing that they are representative of the population of teachers in terms of unobservable characteristics—the key advantage that random probability sampling is supposed to bring.

The overall response rate of the TALIS studies conducted in England are notably larger—if still not perfect. In the [Random probability surveys of teachers](#) section we discussed the minimum response rate criteria—set internationally, and independent of government—for a country to be included in TALIS. The TALIS data for England just about met these criteria. This in turn means that, while there is still likely to be some non-random participation in these studies, the risk of bias in the estimated population parameters (e.g., average working hours) is lower. In other words, the implicit assumption that the TALIS sample is likely to be similar to the population of teachers in unobservable (as well as observable) ways is more credible than for most other teacher surveys in England.

TABLE 2 Recent national probability samples of teachers conducted in England.

| Study | Design | Cost | Response rate (%) | | | Sample size | | | Number of questions |
|--|----------------------|----------|-------------------|---------|---------|-------------|----------|-------|---------------------|
| | | | School | Teacher | Overall | Schools | Teachers | | |
| TALIS 2013 (secondary) | Cluster | £300,000 | 56 | 83 | 47 | 115 | 2496 | 51 | |
| TALIS 2018 (primary) | Cluster | £573,000 | 64 | 85 | 54 | 126 | 2009 | 65 | |
| TALIS 2018 (secondary) | Cluster | | 68 | 84 | 57 | 130 | 2376 | 65 | |
| Working Lives of Teachers and Leaders. Wave 1. | Simple random sample | £2.8m | – | 13 | 13 | – | 9094 | 35min | |
| Ofsted teacher wellbeing survey | Cluster | – | 29 | ~40 | ~12 | 290 | 2293 | ~100 | |
| Teacher Workload Survey 2016 | Cluster | £108,000 | 24 | 34 | 8 | 218 | 3186 | 32 | |
| Teacher Workload Survey 2019 | Cluster | £100,000 | 27 | 40 | 11 | 405 | 7500 | 33 | |

Note: School and overall response rates for the three TALIS samples are reported before replacement schools have been considered. This is to ensure that the information reported in this table is as comparable across studies as possible. The response rates when replacement schools are included is higher (above 75%), thus meeting the OECD's inclusion criteria. For instance, in TALIS 2018, the OECD reports the overall response rate for England to be 73% (primary) and 68% (secondary) respectively. School level information reported for TALIS is hence always before replacement schools have been considered, while the number of teachers is reported in terms of the final sample. Ofsted did not report teacher response rate. From government records, the average school has 19 full-time equivalent teachers, meaning there would be around 5500 teachers within the 290 participating schools. From this, we estimate the teacher response rate to be around 40%.

Why were response rates higher in TALS than other studies? It is likely several factors were at play. The response rate targets—and high-stakes consequences for failing to meet them (effectively being excluded from the study)—are likely to have focused the efforts of those conducting the fieldwork to achieve this goal. As [Table 2](#) illustrates, the project had a realistic budget for a single cross-sectional study of its size and scope, whereas others (such as the 2016 and 2019 workload surveys) were clearly under resourced. All the major teaching unions showed strong public support for TALIS and encouraged their members to participate. Indeed, the major teaching unions in England jointly wrote to the Department for Education to complain about the government's decision to not participate in the 2024 round of TALIS, such was their backing of the study. Yet government strongly encouraged teachers to participate in TALIS 2018 as well, including a letter being written from ministers to schools to make clear how important it was for them to take part. This, together, illustrates the type of backing that random probability surveys of teachers requires to be successful; if government departments and other organisations are going to use this study design, then they need to fully commit to it, providing the resource (and backing) that it needs.

As noted in the [Random probability surveys of teachers](#) section, despite the lower than desired response rates to random probability surveys of teachers, it is still possible to assess their representativeness of the broader population in terms of the characteristics observable in both. The more characteristics this can be done for the better, particularly in terms of those that are likely to be strongly associated with the attributes the survey is attempting to measure. [Table 3](#) thus summarises what each of the recent random probability surveys of teachers in England have done in this regard.

Overall, comparisons between the sample achieved and the population (or, analogously, between sample participants and non-participants) has been very limited. This is likely due—at least in part—to the limited information available about the population of teachers that can be compared to the sample. For instance, despite its very low response rate, the Working Lives of Teachers and Leaders only drew comparisons between the sample and population of teachers in terms of gender, ethnicity and job role. This provides little insight into whether there has, for instance, been selection into the study based on workload (it is an ‘unobservable factor’) despite clear reasons to suspect this might be the case (teachers under workload pressures having less time to complete the survey). Such selection would be particularly problematic, given the importance attached to workload in recent education policy debates—the sample is unlikely to provide unbiased estimates of one of the key issues it was designed to measure. The reality is that claims the data are ‘representative’ (IFF Research, [2023a](#))—and that treating it as a truly random sample—are based on very thin evidence indeed.

More generally, across all studies, checks on the representativeness of the samples in terms of teacher-level variables is particularly scant. This is notable, given how teachers are the primary unit of interest in these studies. Thus, overall, what one can say about the ‘representativeness’ of national probability surveys of teachers in terms of observable characteristics is only very weak, given the extremely limited set of teacher-level attributes most existing studies have considered.

Teacher panels in England

[Table 4](#) turns to the four major teacher panel studies in England, providing some key information about each. It is immediately clear that not all teacher panels are equal; they vary in size, number of questions, frequency and cost. Three are essentially commercial enterprises where researchers or organisations can pay for questions to be asked to teachers on

TABLE 3 Comparisons made between the observable characteristics of random probability samples to the population of teachers.

| | School factors | Teacher factors |
|---|--|---|
| TALIS 2013 (Key Stage 3 teachers) | Ofsted rating, independent/state school, region, size, national examination achievement | Gender, age, subject |
| TALIS 2018 (Key Stage 2 teachers) | School type, Ofsted rating, school performance measures, pupil absence rates, average teacher pay, pupil-teacher ratio, % eligible for FSM | None |
| TALIS 2018 (Key Stage 3 teachers) | Phase (primary / secondary) | Job role, gender, ethnicity |
| Working Lives of Teachers and Leaders. Wave 1. | | |
| Ofsted teacher wellbeing survey | None | None |
| Teacher Workload Survey 2016 | None | None |
| Teacher Workload Survey 2019 | Size, phase, rural/urban, Ofsted rating, region, school type, % eligible for free school meals, % English as an additional language | Age, gender, fixed-term contract, job role, part-time |

Note: Refers to comparisons made either between the sample and population values, or between respondents and non-respondents to the survey.

TABLE 4 An overview of teacher panel surveys in England.

| Name | Respondents | Frequency | Number of questions | Approximate cost |
|--------------------------|-------------|------------------|---------------------|--------------------------|
| TeacherTapp | ~6000–8000 | Daily | 3 questions per day | ~£1000 per question |
| Teacher Voice | ~1000 | 3 times per year | Unknown | £1380–£2000 per question |
| YouGov | ~1000 | Bespoke | Bespoke | Unknown |
| School and College Panel | ~3500 | 6 times per year | 5–7 min each time | N/A (Run by government) |

Note: Number of respondents for the School and College Panel based upon figures for leaders (1447) and teachers (1938) reported in IFF Research (2023c).

the panel, or commission an entire bespoke survey to be conducted. The exception is the School and College Panel (SCP), which is commissioned by the Department for Education for their own use. In the discussion that follows, we focus on TeacherTapp and SCP as the two largest teacher panels currently in operation in England.

First, what do we know about the representativeness of these studies in terms of observable characteristics (noting, as discussed in the [Teacher panels](#) section, that teacher panels will always struggle to make a convincing case they are representative in terms of unobservables)? TeacherTapp have compared their panel to the population of teachers and report that they 'are able to show that our weighted sample mirrors the population of teachers by other characteristics such as Ofsted rating, school FSM %, school governance' (TeacherTapp, 2022). Interestingly, TeacherTapp have also asked their panel several questions that were included in the 2018 TALIS survey—the closest resource England currently has to a genuinely random sample of teachers. They report that they 'check we can replicate key findings from the TALIS questionnaire, which is the closest we've got to a true random sample in England'.

The SCP also report how their panel compare to the broader population of teachers. To illustrate this, [Table 5](#) provides comparisons between the SCP and population values, and between respondents to the first wave of the Working Lives of Teachers and Leaders (WLTL)—the most recent random probability survey of teachers in England—and the sample originally drawn. The bottom row presents the response rate for both surveys, illustrating that these are rather similar (13% for the WLTL versus around 9% for the SCP). Both present evidence for just a handful of variables, with data from the SCP appearing just as similar to the population values reported as the WLTL (the random probability survey)—though with the obvious caveat that different sets of variables across the two studies are being compared. Nevertheless, this again points towards there being little firm evidence that random probability samples of teachers with low responses rates are any superior to teacher panels in terms of their representativeness (in terms of a very limited number of attributes that can be observed).

Turning to other issues, size is often an important consideration in survey design, such as for producing separate results for sub-groups (e.g., primary versus secondary teachers). The TeacherTapp panel achieves sample sizes similar or larger than recent random probability samples conducted in England (~8000–10,000), while those from the other panels are notably smaller (~1000 for Teacher Voice and many of the You Gov studies conducted with teachers). Nevertheless, it is clear that at least some of the Teacher Panel data collected in England is of sufficient size to explore differences across sub-groups within the sample collected.

Focusing on cost, most random probability surveys of teachers tend to be quite short to encourage high response rates. As illustrated in [Table 2](#), a standard size is around 50 questions. The approximate cost of asking 50 questions via TeacherTapp is £50,000. In comparison, the 2016 and 2019 workload surveys—which included around 30 questions—cost around twice as much (£100,000). Although questions from other teacher panels appear more expensive (e.g., up to £2000 per question in Teacher Voice), it is clear that the cost of data collection is no more expensive, and probably cheaper, than conducting a random probability survey.

In terms of timeliness and flexibility, teacher panels are clearly superior. Whereas national probability samples typically occur just once per year, [Table 5](#) illustrates how teacher panels provide data much more frequently. TeacherTapp is an extreme example, where teachers answer a small number of questions (3) every day. This particular teacher panel can thus gather information from teachers in a very timely manner—reacting to events (e.g., policy announcements) very quickly.

TABLE 5 Comparison of the School and College Panel (SCP) and Working Lives of Teachers and Leaders (WLTL) to teacher population characteristics.

| (a) Working lives of teachers and leaders (random probability survey) | | |
|--|------------------------------|-----------------------|
| | Initially sampled (%) | Responding (%) |
| Job role | | |
| Heads | 6 | 7 |
| Assistant heads | 5 | 5 |
| Deputy heads | 3 | 4 |
| Leading practitioner | 1 | 1 |
| Class teacher (not ECT) | 71 | 71 |
| ECT year 2 | 10 | 11 |
| ECT year 1 | 3 | 2 |
| Phase | | |
| Primary | 48 | 51 |
| Secondary | 47 | 44 |
| Other | 5 | 5 |
| Gender | | |
| Male | 25 | 24 |
| Female | 75 | 76 |
| Other | 0 | 0 |
| Ethnicity | | |
| Asian | 5 | 3 |
| Black | 2 | 2 |
| Mixed | 1 | 1 |
| White | 83 | 86 |
| Other | 1 | 0 |
| Unknown | 8 | 7 |
| Total | 88,470 | 11,177 |
| (b) School and college panel (teacher panel) | | |
| | Population (%) | Panel (%) |
| Phase | | |
| Primary | 52 | 49 |
| Secondary | 48 | 51 |
| Academy status | | |
| Academy | 47 | 60 |
| Not academy | 53 | 40 |
| Free school meal quintile | | |
| Quintile 1 | 11 | 10 |
| Quintile 2 | 15 | 15 |
| Quintile 3 | 19 | 20 |
| Quintile 4 | 22 | 25 |
| Quintile 5 | 32 | 31 |

(Continues)

TABLE 5 (Continued)

| (b) School and college panel (teacher panel) | | |
|---|-----------------------|------------------|
| | Population (%) | Panel (%) |
| Ofsted rating | | |
| Outstanding | 12 | 12 |
| Good | 64 | 62 |
| RI | 7 | 9 |
| Serious weakness | <1 | <1 |
| Special measures | <1 | <1 |
| Unknown | 15 | 16 |
| Response rate | 9 | |

Note: Figures for WLTL are taken from table 3.4 of IFF Research (2023a). Figures for SCP are taken from tables 7–10 of IFF Research (2023b). For the SCP we report figures for teachers, excluding leaders, wherever possible. We do not report information on job role, as the panel was purposefully designed to over-represent school leaders. Response rate for SCP is based upon table 5 of IFF Research (2023b).

Another advantage of teacher panels is that they facilitate tracking the views of the same group of teachers over time. This is much more challenging and expensive to achieve in most random probability surveys of teachers. Indeed, although the WLTL has been designed as a longitudinal study (with a budget of almost £3,000,000 for five waves), the 13% response rate to the baseline survey means this has essentially ended up as a convenience sample. As this experience shows, obtaining longitudinal data from a genuinely random sample of teachers over time has proven to be almost impossible to achieve in England (even when the budget has been large). Arguably, teacher panels are hence also the optimal choice for any data collection that involves surveying the same teachers over time.

Finally, what about the amount of data available? As noted above, most random surveys of teachers contain around 50 questions in a single academic year (see Table 2). In contrast, knowledge about teachers involved in panel studies accumulates over time. Take Teacher-Tapp, for example. If a teacher completes all three questions they are asked each day for one year, then there will be over 1000 pieces of information available about them. Such a vast array of data can provide incredibly rich insights into teachers' lives, including how their views of certain issues change over time and correlate with other factors of interest. Random probability samples are unlikely to ever be able to provide the same level of detail about teachers and the teaching profession.

CONCLUSIONS

Teachers are one of the most important inputs into the education of young people. This has led to widespread interest into their views, attitudes and opinions about the education system and, in particular, their job. To gather information from teachers, governments often attempt to collect data from a random probability sample, following a textbook understanding that this is the 'gold standard' approach. Yet data are also increasingly being used from teacher panel surveys to provide insights into the teaching profession. Although in theory such data may be less representative of the broader population of teachers—as compared to a true random sample—these have important advantages in terms of cost, speed and flexibility, while also allowing one to track the views of the same group of teachers over time.

The main aim of this paper has been to review the theoretical advantages and disadvantages of these two approaches to surveying teachers, before discussing how this plays out

in practice via a case study of England. We discuss how it is the randomness of the sample that provides a very high degree of confidence that the data from random probability samples will truly be representative of the broader population of teachers in both observed and unobserved ways. This then allows one to generalise statements based upon the sample to the population of teachers as a whole (e.g., the average number of hours teachers in England work per week).

However, in reality, all that glistens is not gold. In practice, the randomness of national probability surveys of teachers is often severely undermined by their anaemic response rates. In England, often only around one-in-ten of those teachers initially randomly selected goes on to complete the questionnaire. Hence the major theoretical advantage of random probability surveys over teacher panels is very likely lost. The low response rates also means that the implicit assumption that the sample is representative of the broader population of teachers in terms of unobserved variables is no longer credible. Moreover, there is little hard evidence that random probability samples with low response rates are any more representative of the population of teachers in terms of observable characteristics than teacher panels either. Thus, the problematic execution of recent random probability surveys of teachers in England has simply led to slower, less rich, more costly and less nimble data about the teaching profession than if collection were done via teacher panels instead.

These observations have important implications. In England, the Department for Education's strategy to generate evidence about teachers and the teaching profession must be revised. Random probability surveys should continue to play an important role. But, when these are conducted, there needs to be much greater commitment to executing them properly. This, critically, means doing everything possible to obtain a high response rate. The TALIS data collections have shown how achieving reasonably high response rates from a random probability sample of teachers in England is possible, but only when there is real resource and energy devoted to achieving it. Recently, there have been too many half-hearted attempts to gather data from a random sample of teachers that have ultimately resulted in failure.

How should such a study be designed? One option would be for England to rejoin TALIS in future waves, given the reasonable degree of success this study had previously (and clear union backing for it). Alternatively, a biennial study could be conducted with a simple random sample of 2000 teachers (1000 primary and 1000 secondary).¹⁰ The survey could be of moderate length (e.g., 30 min) to encourage participation and, critically, offer a significant financial incentive (e.g., £50–£100 for each responding teacher—equivalent to paying them £100 to £200 an hour). With a further £100,000 allocated to cover fieldwork expenses, the annual cost of such a survey would be around £300,000. Its cost would hence be in a similar ballpark to TALIS and would represent just 0.0005% of the £57 billion annual grant allocation to schools (School Funding Statistics, 2023). This is likely to give England the best chance of generating truly representative and high-quality data about teachers and the teaching profession, while also providing an important additional yardstick of observable characteristics that teacher panels could benchmark their data against.

In terms of teacher panels, our advice is that further details are published on how they have checked the representativeness of their data—for example, how their respondents compare to known characteristics of the teacher population. Such information could then be updated annually, as the composition of their panels change. Similarly, if/when higher quality data from a truly random sample of teachers in England is next collected, teacher panels could publish details on how their samples compare to these resources in terms of the questions asked (as TeacherTapp has previously done with TALIS). More generally, teacher panels clearly have an important ongoing role to play in key education policy debates. In many ways, they are the only route to obtaining timely information about topical issues and of tracking a group of teachers over time (including how their workloads, thoughts

and feelings vary over the course of an academic year). Although their samples may not be random, neither are those from most of the available alternatives either.

FUNDING INFORMATION

No funding to declare.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ETHICS STATEMENT

The BERA code of ethical practice has been followed.

ENDNOTES

- ¹ Occupational codes are usually assigned in general population surveys after respondents have self-reported the job that they hold.
- ² Or, possibly, all teachers within each selected school asked to participate.
- ³ A key limitation with using simple random sampling (rather than cluster sampling) is that only one or two teachers are likely to be selected from each school, offering no possibility to link the views of teachers to the views and actions of school leaders. This in turn makes simple random sampling less useful for understanding how teacher's views are impacted by workplace environmental factors, such as school leadership.
- ⁴ In other words, if it were possible to conduct the same survey on the same population many times, and then take the average across these surveys.
- ⁵ The TALIS survey design includes 'replacement schools' in the response rate. Essentially, if a school refuses to take part, another school (that is adjacent on the sampling frame) is allowed to take its place. This is essentially a form of imputation. The minimum criteria the OECD set for TALIS is that the before replacement school response rate should be at least 50%, and the after replacement school response rate should be 75%.
- ⁶ Or, analogously, how the final achieved sample compares to the initially randomly drawn sample in terms of observable characteristics.
- ⁷ In theory, if non-response is completely random, then the sample will still be very similar to the population in unobservable (as well as observable) ways. The issue, however, is that there is no way to establish whether non-response is random in terms of unobservable characteristics. Balance across the sample and population in terms of observables only provides reassurance on this point to the extent that these characteristics considered are correlated with potentially important unobservable characteristics.
- ⁸ It is, however, possible to compare the characteristics of teacher panel participants against the broader population of teachers and reweight the sample accordingly. This is essentially the same approach as discussed for random probability surveys that suffer low response rates.
- ⁹ Despite this, it is common practice that such inferential statistics are reported anyway, ignoring the fact that the sample has not been randomly selected.
- ¹⁰ A minimum sample size of 1000 is often used in polling, with this giving an approximate margin of error of $\pm 3\%$ from the sample to the true population value.

REFERENCES

- EDAPT. (2023). *Teacher Strikes 2023 (Easter Update)—What happens now? – Edapt*. <https://www.edapt.org.uk/support/knowledge-base/teacher-strikes-2023-easter-update/>
- Gorard, S. (2015). Rethinking 'quantitative' methods and the development of new researchers. *Review of Education*, 3, 72–96. <https://doi.org/10.1002/rev3.3041>
- Hanushek, E. (2011). Valuing teachers: How much is a good teacher worth? *Education Next*, 11(3), 40–45.
- IFF Research. (2023a). *Working lives of teachers and leaders—wave 1 Technical report*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1148572/Working_lives_of_teachers_and_leaders_-_wave_1_-_technical_report.pdf

- IFF Research. (2023b). *School and College Panel Technical report: 2022 to 2023 academic year*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1137840/School_and_College_Panel_technical_report_autumn_2022.pdf
- IFF Research. (2023c). *School and college panel—November 2022*. Department for Education Research Report. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1137536/School_and_college_panel_November_report.pdf
- Jerrim, J. (2021). PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK? *Review of Education*, 9, e3270. <https://doi.org/10.1002/rev3.3270>
- Jerrim, J., Allen, R., & Sims, S. (2022). *How did the COVID-19 pandemic affect the wellbeing of teachers at work?*
- Jerrim, J., Sims, S., Taylor, H., & Allen, R. (2020). How does the mental health and wellbeing of teachers compare to other professions? Evidence from eleven survey datasets. *Review of Education*, 8, 659–689. <https://doi.org/10.1002/rev3.3228>
- OECD. (2005). *Teachers matter. Attracting, developing and retaining effective teachers*. OECD Publishing.
- School Funding Statistics. (2023). *Financial year 2022–23*. <https://explore-education-statistics.service.gov.uk/find-statistics/school-funding-statistics>
- TeacherTapp. (2022). *How it works—TeacherTapp*. <https://teachertapp.co.uk/how-it-works/#representative-data>
- Worth, J., & Van den Brande, J. (2019). *Teacher labour market in England annual report 2019*. NFER.

How to cite this article: Jerrim, J. (2023). Teacher surveys: The pros and cons of random probability surveys versus teacher panels. *Review of Education*, 11, e3428. <https://doi.org/10.1002/rev3.3428>