



Department  
for Education

# Teacher Workload Survey 2016

Technical report

February 2017

John Highton, Sarah Leonardi, Neil Richards and Arifa  
Choudhury, CFE Research

Dr Nicholas Sofroniou and Dr David Owen, Institute for  
Employment Research, University of Warwick



Social Science in Government

# Contents

Annex 1: Survey design	4
Background to the survey design	4
Setting a baseline describing teachers' workload	4
Comparability with past surveys	4
Rationale for using a survey rather than a diary	5
Design timeline	8
Annex 2: Sampling	10
Design considerations	10
Comparability with TALIS and prior diaries	10
Prior diary studies	11
The practicalities of research in schools	11
Sampling approach	12
School selection probability	12
Sampling	13
Annex 3: Cognitive testing	15
Selection of schools	15
Testing process	16
Feedback from testing	16
Changes made as part of the questionnaire development phase	16
Annex 4: Final questionnaire	21
Working hours questions – TEACHER route	21
Working hours questions – HEADTEACHER route	27
Perceptions and drivers of workload	31
Annex 5: Response rates	42
School cooperation rate	42
Teachers' response rate	43
Improving cooperation and response rates in subsequent surveys	43
Annex 6: Data processing and tabulation	45
Cleaning	45
Data Matching	46

Processing and tabulating	46
Annex 7: Weighting	48
Annex 8: Modelling the factors which impact on teacher workload	49
Introduction to multilevel models	49
Between- and within-school variance components	49
Procedure used for multilevel modelling	50
Explanatory variables evaluated as separate models	53
The final models of total workload	57
Contributions of the explanatory variables to fitted scores	60
References	63

## Annex 1: Survey design

This section describes the survey design process for the Teacher Workload Survey. The section summarises the key design decisions and rationale for the final structure of the survey instrument.

The survey design was led by a steering group, which was comprised of DfE staff and representation from each of the main teaching unions.

### Background to the survey design

#### Setting a baseline describing teachers' workload

The invitation to tender issued by DfE for this survey stated that a robust baseline measure of teachers' workload was required for primary and secondary schools in England. Existing data either described a limited sub-set of the teacher population robustly, or collected less robust data over the whole population. As a result, it was difficult for DfE and other parties interested in teachers' workload to draw comparative conclusions about working hours. A survey was required to address this gap, set a baseline describing workload and, in subsequent years, allow DfE to monitor changes in workload comfortable in the fact that differences between years were meaningful.

#### Comparability with past surveys

The Teacher Workload Survey was designed to combine elements from two other survey instruments: the Teaching and Learning International Survey (TALIS), which is administered to OECD countries, and past teacher workload diaries commissioned by DfE.

**The principle challenge in the design was to achieve as close comparability with TALIS as possible whilst recognising that the level of detail on workload issues in TALIS was limited compared to past diary studies.**

CFE designed an online survey of ten minutes in length in order to minimise the research burden faced by an audience whose time is already limited. In designing the questionnaire, the key questions pertaining to workload within TALIS were identified, alongside redesigning questions from previous diary studies (last commissioned in 2013). As comparability with TALIS was central to the design, the wording of these questions and, where possible, their order was retained. However, some changes were necessary for a variety of reasons including the relevance and grammatical structure of TALIS questions in relation to an English audience. More detail on this is covered within Annex 3: Cognitive testing.

Items describing workload in prior diary studies were also important to include. In particular, past diaries recorded detail on professional activities undertaken by teachers that featured within the DfE Workload Challenge. Understanding these activities in more detail and designing a study that allows DfE to confidently measure changes in the amount of time spent on these activities in the future was therefore an important design consideration.

The relationship between TALIS and the diary studies is mapped in Table A1.1 at the end of this section. The relationship between individual survey items was an important consideration for the study and results in some conflicts between each. For example, TALIS records the time spent on a range of pupil guidance activities as “*students counselling (including student supervision, virtual counselling, career guidance and delinquency guidance)*”. The equivalent measures in prior diary studies account for a much greater slice of teachers’ time and, importantly, separate out activities that are important in an English context, such as pupil discipline. From a cognitive perspective, this item also includes use of the word *student* rather than the preferred noun of *pupil* in the English context. This is a good example of some of the wording changes made based on cognitive testing.

## Rationale for using a survey rather than a diary

Past diary studies had poor response rates from teachers (TNS-BMRB, 2014). Response rates are governed by a range of factors, which were considered in the overall design of the survey instrument and its administration. These factors can be classed under categories that include:

1. **Research Burden for the teacher.** The last diary study asked teachers to code their time in five-minute intervals for an entire, randomly selected day within a given working week. The practicality of this approach in relation to a teacher’s working day is questionable and it was theorised that teachers would mainly complete their diaries after their working day. By nature, this would lead to some level of estimation and compare unfavourably to those completing the survey during the day.

In addition, the diary design during its last iteration was complex and used a paper-based method of data collection. Whilst a paper-based approach is appropriate for a diary, the burden for the teacher on the given collection day *given the topic of the study* was deemed inappropriate. Reducing the survey length is one of several standard design measures relating to response rates (see Krosnick, 1991).

2. **Research burden for the school.** The sampling method for the diary study used was a traditional two-stage probability proportional to size (PPS), as described in Annex 2. In short, this method involved firstly selecting schools stratified by size and other factors as the primary sampling unit (PSU), then randomly selecting a set number of teachers within a school to take part in the survey. The second part of this approach is potentially

problematic for the school without considerable support from the research contractor and DfE because:

- The method requires random selection of a set number of teachers. The individual within each school responsible for distributing the diary needs to apply a consistent and potentially complex randomised approach (*1-in-n* selection from a random seed, Kish grid, etc.). There is a significant time commitment from the school to deliver this and to administer surveys to individually selected teachers. There is a dissonance between this time requirement, the wish to reduce burden on participating schools and the nature of the workload study.
- There is also a significant time commitment required from the contractor to deliver this approach, especially given the number of PSUs (i.e. schools) involved in the survey. This has a major cost implication, which requires a resulting and consummate benefit to justify.

Taken together, these factors make it less likely that schools would agree to take part in a study, which has a subsequent impact on cooperation and response rates from schools. However, it is important to note that the two-stage process is a methodologically more rigorous approach to take to the study, because it corrects for the bias introduced through sampling by the size of the PSU (that teachers in larger schools are more likely to be selected to take part in the study).

3. **Design and administration of paper responses.** It is not possible to use validation in paper surveys. This can lead to teachers placing their own interpretation on a question and amending it to fit what they think it should say. This invalidates the response. All possible routes through a survey also need to be shown on a paper document. This means the survey itself may appear longer than it is, and lead to errors in the way an individual respondent completes the survey, including completion of incorrect questions for that individual (based on their role for example). All of these issues can be managed by a survey script and address non-response through invalid completions.
4. **Requirement to post returned diaries.** Once completed, diaries needed to be posted back by teachers. This introduced a further step in the process that was outside the control of the contractor. Return envelopes could be lost, diaries completed but unreturned, or the perception of burden increased in the mind of the teacher.
5. **Contact and reminders to teachers.** In past studies, paper diaries were passed to selected teachers via the school. Subsequently, it was a burden on the school to also ask them to send reminders to the selected teachers. As such, it was not simple to administer reminders using the paper-based diary method to a random selection of teachers. The census method via a school's internal email system adopted for the Teacher Workload

Survey addressed this problem and made it simple for school administrators to send reminders.

6. **Timing.** In 2013, TALIS was administered in the weeks leading up to Easter. It was important that the workload survey replicate this timing. This led to some one-off time pressures this year due to the date of commissioning; it is assumed that subsequent workload surveys will be commissioned earlier to address this issue and allow more time for planning. However, this study also had a relatively short fieldwork window of three weeks and this duration is likely to remain fixed for two main reasons. Firstly, it will be important to at least repeat fieldwork in the same half-term of the school year. This by nature limits the fieldwork period. Secondly, it is also important that fieldwork is conducted to minimise any differences that might arise between weeks. The first and last week of a term can be atypical of others as pupils return or prepare for a school holiday. This further limits the available fieldwork time.

An additional factor outside of response rate considerations favouring a survey approach was the comparability with TALIS (2013). The TALIS survey used the last full working week of the teacher as the unit measurement rather than a random working day. However, it is important to note that both TALIS in 2013 and past diary surveys used the traditional two-stage PPS approach. In this respect, the Teacher Workload Survey does differ in design. The implications for comparability are minimal as outlined in Annex 2: Sampling.

## Design timeline

Once all the factors above were taken into consideration, the survey was designed in January 2016 for initial testing in February 2016 (including the cognitive testing process, detailed in Annex 3: Cognitive testing).

**Table A1.1: Comparison between TALIS 2013 and Teacher workload diary questions**

Diary workload measures	TALIS 2013 workload measures	Average hours spent (secondary phase)		
		Diary, 2013	Diary, 2010	TALIS, 2013
Regular timetabled teaching	Of this total, how many 60-minute hours did you spend on teaching during your most recent complete calendar week?	19.6	18.5	19.6
Non-regular teaching				
Other teaching activity				
Covering*				
Coaching/rehearsing	Engaging in extracurricular activities (e.g. sports and cultural activities after school)	7.8	6.9	2.2
Supervising pupils	Student counselling (including student supervision, virtual counselling, career guidance and delinquency guidance)			1.7
Registration/pastoral				
Discipline/praising pupils				
Contact with parents	Communication and co-operation with parents or guardians			1.6
Other non-teaching pupil/parent contact	<i>Would be one of the two listed above</i>	<i>n/a</i>		
Planning/preparing lessons	Individual planning or preparation of lessons either at school or out of school	18.9	16.2	7.8
Other preparation/assessment				
Assessing/marking	Marking/correcting of student work			6.1
General staff/management meetings	Team work and dialogue with colleagues within this school	4.6	2.6	3.3
Other contact/interaction with staff				
Appraising/coaching/mentoring others	Participation in school management			2.2
School policy development and financial planning				
Contact with someone outside of school				
Other school/staff management				
Keeping records		2.3	2.1	4.0

Diary workload measures	TALIS 2013 workload measures	Average hours spent (secondary phase)		
		Diary, 2013	Diary, 2010	TALIS, 2013
Organising resources and premises	General administrative work (including communication, paperwork and other clerical duties you undertake in your job as a teacher)			
Other kinds of administrative activities				
Training or development activity	<i>No specific measure in TALIS, would be under "other"</i>	1.7	3.1	2.3
Other professional development				
Non-working breaks/absences during school hours	<i>Not collected in TALIS</i>	0.8	0.5	n/a
Own time outside school hours				

\* Classed as non-teaching contact in diary, not teaching

## Annex 2: Sampling

### Design considerations

The sampling design needed to address several competing issues:

1. Retain comparability with TALIS (2013) and, to a lesser extent, earlier diary studies;
2. Obtain a robust achieved sample of the target audience and critical sub-groups with a better response rate than the prior diary study; and
3. Do so in a cost-effective way.

### Comparability with TALIS and prior diaries

TALIS used a clustered two-stage sampling method utilising a Probability Proportional to Size (PPS) approach. The two stages of PPS are traditionally:

- To stratify the primary sampling unit (PSU) firstly by some measure of size, then by other strata of interest. In this case, the PSU was the school and the measure of size was the number of teachers. In the case of English schools in TALIS, the stratification variables were by: funding type (Independent v State funded), two size bands and four regions (OECD, 2014, p.80).
- The second stage is to then randomly select the same number of participants within each selected school to take part in the study.

For English schools in TALIS in 2013, *lower secondary* schools were in scope (the primary and upper secondary phases were out of scope). The study achieved excellent response rates of 75% for schools and 83% for teachers (Micklewright *et al*, 2014, p.15). However, achieving this response rate required a significant amount of support and administrative work on behalf of the research contractors including liaison and support activity with designated in-school coordinators, the production and distribution of manuals covering the sampling approach and in-school administration, other training materials including videos and the use of bespoke communication methods and national support infrastructure (Micklewright *et al*, 2014, pp. 85-95). Minimum school participation rates and teacher response rates also formed part of the contract and the necessary budget was provided to achieve this. The extent to which the same level of resource was justified for this study was arguable. Furthermore, the time-frame allowed to deliver the survey was a fraction of that allowed for TALIS, which meant the same level of support and survey administration resourcing was impractical for the Teacher Workload Survey.

## Prior diary studies

The last diary study was administered in 2013. It used a random sampling approach by selecting teachers to undertake the survey using the School Workforce Census (SWC<sup>1</sup>) as a sampling frame. Theoretically, this approach removes much of the selection bias from using a two-stage clustering method. However, the response rate for the 2013 diary survey was only 15% leading to significant non-response bias, especially as there was also differential response rates by teacher subgroup (TNS-BMRB, 2014).

Paper was the only viable survey mode using SWC as the sample frame for two reasons. Firstly, SWC does not hold school email addresses for teachers and hence was unsuitable to sample and contact teachers directly for an online study. Secondly, whilst a school telephone number was recorded, it was impractical and burdensome to repeatedly call schools in order to interview individual teachers working in them.

Given these constraints and the deviation in this approach from TALIS, an adapted PPS approach more in line with the TALIS method was agreed.

## The practicalities of research in schools

The TALIS technical report noted that differences in the structure of the teaching body were greater between schools than within schools (OECD, 2014). As a result, the key element of any sampling approach was to ensure that there was a representative sample of enough schools in the sample frame from which teachers could be drawn to minimise clustering effects.

A differential response from teachers was experienced in the last wave of TALIS and between roles and phases in the 2013 Teacher Workload Diary Survey. This was addressed by weighting for non-response. Weighting was an essential element of this study as the sampling method used (PPS without adjustment) delivered an unbalanced sample. The traditional second phase of a PPS method would then see a set number of teachers in the selected schools randomly sampled to account for a higher selection probability of teachers in larger schools. Weighting was, therefore, essential to adjust for this bias, and to account for non-response from teachers.

Some support and guidance activity with schools was undertaken in the current research to maximise response rates. This included engaging with schools during the recruitment phase and providing a large amount of explanatory materials to gain support and buy-in prior to the fieldwork going live. Help from members of the DfE team was especially helpful and effective in

---

<sup>1</sup> DfE state that: "The school workforce census is a statutory data collection that takes place each autumn... [SWC] collects data on all teaching and support staff in regular employment."  
<https://www.gov.uk/guidance/school-workforce-census>

encouraging a number of schools to participate in the study. Once in field, chasing activity via email and phone to schools was undertaken and reminder emails drafted for school coordinators to send to their teaching staff.

## Sampling approach

The overall sampling approach used the first stage of a Probability Proportional to Size (PPS) method to sample schools. Specifically, schools were selected by the number of teaching staff listed in the 2014 School Workforce Census (the most recent available at that time). Sampling was disproportionate by school phase. Secondary schools are larger cluster units than primary, so the design was such to ensure a large enough number of secondary clusters were present. Without oversampling, there was a significant chance of bias at the secondary school level, because too few PSUs would have been selected.

900 schools were originally sampled. 600 formed the initial sample and 300 were held in reserve. One school was removed almost immediately after drawing the initial sample (due to missing contact details listed in the SWC) and never contacted and a matched school was drawn from the reserve to replace it. All of the remaining 899 schools were eventually invited to take part via email at some point in the sampling phase. Of these, 757 were followed up using telephone methods.

## School selection probability

The selection probability of the school (PSU) was calculated using the following formula:

$$\frac{(\text{Population of the cluster (eligible teaching staff in school)} * \text{Number of clusters sampled})}{\text{Total population}} \quad (e1)$$

There are two ways to conduct this calculation: 1) At the total school population level or; b) by the disproportionately stratified population i.e. for primary, then separately for secondary. Both selection probabilities were calculated.

The second stage of the PPS process usually involves randomly selecting the same number of secondary sampling units (in this case teachers) within each cluster. This means the probability of selection for each teacher within the full sample is the same. However, it was agreed that this stage would be removed for this survey for several reasons:

1. The survey concerns workload. Asking schools to draw a sub-sample of teachers on a random basis within their school would add a significant amount of work for the school and would increase the refusal rate without much more guidance and support.

2. The subsequent increase in cooperation rate through simplifying administration would increase the total number of interviews achieved. Note the final number of completed questionnaires for this study was twice that of previous diary studies using the second stage method (and over three times that of the 2013 diary study).
3. Response rates per school would differ meaning non-response weighting would be required. In the TALIS survey, thresholds were set stipulating minimum response rates and significant resource was in place to ensure these thresholds were met. The budget required for this level of support would have been disproportionate for the workload study.
4. The variation in schools sizes (a function of the number of teachers) varies markedly by school. For example, the smallest secondary school sampled employed seven teachers and the largest 171. In the case of primary schools, teacher counts ranged from three to 53. The only way to achieve an equal selection weight by teachers would be to set a very small target per school (three) and sample far more schools. Given the time constraints and value for money, this was unrealistic. In the case of TALIS, only secondary schools were in scope, so the amount of variation in the cluster size was less. Furthermore, there was also some variation in the TALIS cluster size as *“if a school had 30 teachers or less, all were sampled for convenience.”* (Micklewright *et al*, 2014).
5. Disproportionate sampling by phase was also important. Without this, far more primary schools would have been selected than secondary. As a result, most of the secondary teacher responses would have come from a small number of schools, which may have introduced significant bias.

Rather than sample teachers in a second stage of PPS, a census was taken within the randomly selected schools instead. Whilst this addresses all the points above, it does mean that teacher responses within these randomly selected schools were self-selecting and that the sampling method does not correct for the unequal selection probabilities that result from selecting schools by the size of the teaching body. As noted in the survey design rationale (Annex 1), this design is a deviation from the ideal two-stage approach and leads to unequal selection probabilities for teachers which is subsequently corrected through weighting. This is justified by a larger achieved sample size and a much simpler process of administration for schools resulting in improvements in response rates.

## Sampling

The sampling procedure originated from a complete list of all schools in England obtained from Edubase. From this, any non-relevant school types were removed, leaving the following school types:

Academy Converter, Academy Special Converter, Academy Special Sponsor Led, Academy Sponsor Led, Community School, Community Special School, Foundation School, Foundation Special School, Free Schools, Free Schools Special, Studio Schools, Voluntary Aided School, Voluntary Controlled School

Any schools marked as “closed” or “proposed to open” were removed, followed by any schools specified only as “16 plus” or “Nursery” in the Phase field.

Separately, the School Workforce Census (SWC) was used to aggregate the number of teachers currently at every school, at the level of teachers, phase of education, type of establishment, GOR (region) and gender (whether the school was for boys/girls or mixed). The Local Authority Establishment (LAESTAB) number for every school served as a unique identifier to merge this information in with the master dataset.

At this point, any school, where there was no information on number of teachers per school, was given a value of the median number of teachers per school phase.

Data was then split into Primary and Secondary phases and both datasets were sorted on the number of teachers per school, followed by the phase of education, type of establishment, GOR and gender. With the datasets primarily sorted in descending order of number of teachers, an algorithm was devised to calculate the cumulative numbers and percentiles of teachers within each dataset, allowing selection of exactly 175 primary schools equally spaced throughout the primary school dataset and 125 secondary schools equally spaced throughout the dataset, for a total of 300 schools.

The process was repeated twice so that the schools immediately either side of the selected school were also selected and added to separate lists, giving a total of three school lists of 300 schools, each as equal in characteristics as possible, which enabled tracking of the three separate sample groups throughout the recruitment phase, totalling 900 selected schools.

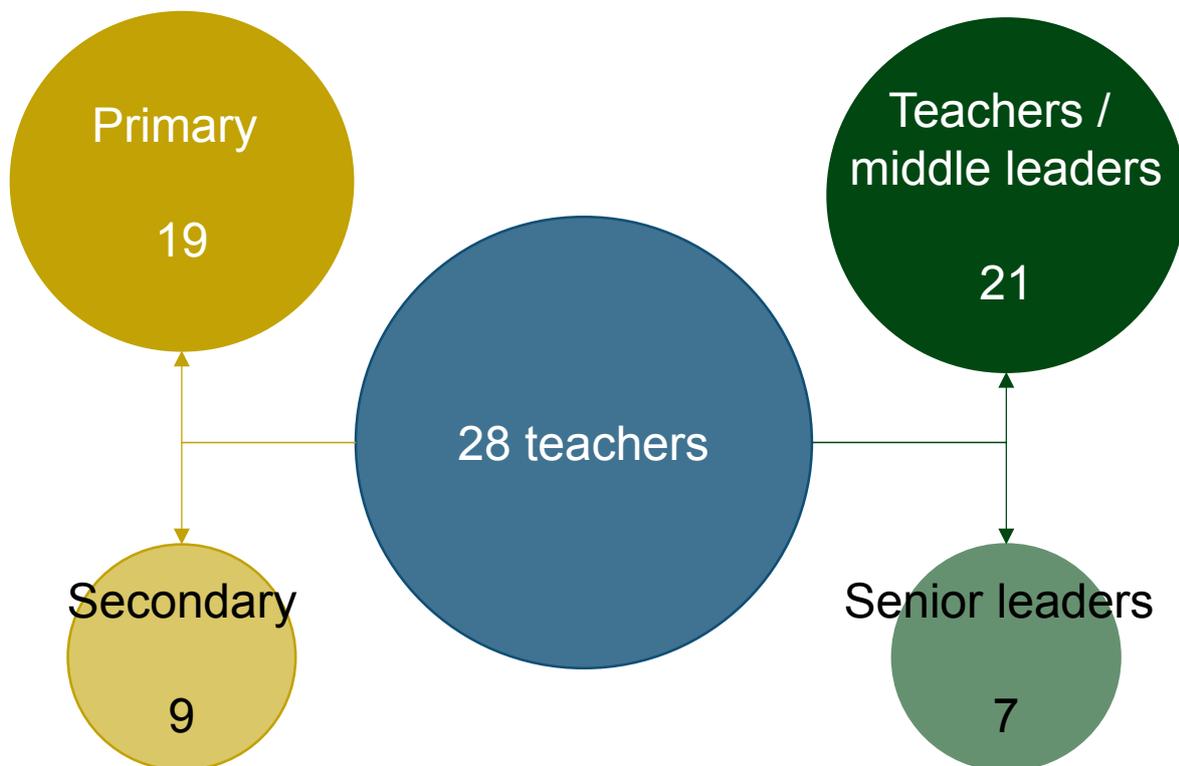
## Annex 3: Cognitive testing

This section describes the process of cognitive testing of the Teacher Workload Survey. The cognitive testing process took place in February 2016 in advance of the live survey, which was conducted in March 2016.

### Selection of schools

Teachers were interviewed in ten schools for the cognitive testing who were not on the survey sample for the mainstage fieldwork. These schools were sampled by factors such as phase (primary and secondary), size and geography. As the survey design included specific questions tailored to a teacher's role/level of responsibility, the initial discussions with each school ensured that there would be good representation of teachers across all levels (headteachers, deputy headteachers and classroom teachers) included in the testing. The testing then took place in ten schools (seven primary and three secondary) across England: four in the Midlands, two in the North East, two in the North West and one in both the South and Yorkshire. The breakdown of interviews is described in figure A3.1 below.

Figure A3.1: Sample split for cognitive testing



CFE sent a Cognitive Testing briefing document (see end of this annex) in advance of initial contact with each school. This summarised the key requirements.

## **Testing process**

The majority of testing took place on-site. A CFE team member attended the school and spent time with the teachers. Group testing usually included between 3-4 teachers. In the case of senior leaders, these interviews comprised entirely of one-to-one interviews with headteachers. Other senior leaders (deputy headteachers) did take part in some of the group work.

CFE's team member initially observed and recorded while the teachers completed the survey online, taking particular note of the length of time taken overall, questions which proved to be stumbling blocks, or any technical issues which prevented satisfactory completion of the survey first time. The remainder of the session was open to share thoughts on the survey and discuss any questions, in terms of wording or comprehension, which teachers felt merited further discussion. While completion of the survey usually stimulated discussion around the subject matter in general, CFE were able to moderate discussion where necessary towards relevant questions and wording within the questionnaire, so that as much relevant feedback as possible was collected and recorded.

The full structure of the cognitive testing session was detailed in a briefing document (shown at the end of this Annex). Each session took 45 minutes to 1 hour depending on the availability of teachers at each establishment, and all sessions were positive and well-received. Many schools were keen for their results to be included in the survey (which was sadly not possible as the questions changed as a result of the testing), but this was a good indication that the surveys were being answered "as live" and were highly salient to teachers.

## **Feedback from testing**

In total, 28 teachers completed a cognitive test and provided a large amount of feedback and suggested improvements. These were collated and reviewed in order to make a series of suggested revisions to the first draft questionnaire. Below is a list of resulting changes that were made between the cognitive testing and live surveys.

## **Changes made as part of the questionnaire development phase**

- The first draft questionnaire was too long, primarily because it included some test questions and variants. The final agreed survey was timed at 15 minutes and the survey introduction, recruitment materials and resulting guidance were adjusted accordingly.

- Part-time teachers particularly felt that reported hours needed to be considered alongside how long they were supposed to work. A question was added to record the number of contracted hours per week.
- Instructions were shortened and simplified where possible as they were deemed confusing in some instances. For example, the instruction on how to record hours to the nearest half hour was revised considerably.
- Teachers spent a long time thinking about each individual listed activity and then totalled the hours to cross-reference with the total. Whilst such cross-checking is potentially desirable, it did lead teachers to spend much longer completing the survey. In addition, some questions (such as questions 6 and 7, Annex 4: Final questionnaire) were designed as sub-categories and more detailed items relating to some of the broader TALIS items such as “General Administration”. As a result, checking totals could lead to error. Further clarification and instruction was included in the survey to reassure teachers on cross-checking. Also note that the TALIS survey reports a discrepancy between the “total” hours question and the addition of the individual units. Removing the “total” question (question 2 in Annex 4: Final questionnaire) was discussed as part of the design process. However, it was agreed that this question should be retained to ensure comparability with TALIS.
- Scales of “a bit too little/too much” were changed to “too little/too much” to reflect the feedback that teachers felt this was hard to quantify.
- “Information meetings” was removed because a few participants were not clear what it referred to, and replaced with “staff meetings” which was mentioned several times.
- Some more administrative tasks were split into a separate question because of the importance of different aspects of administration in the DfE Workload Challenge<sup>2</sup>.
- Wording for some senior leader activities was revised – for example lesson planning was removed from the draft questionnaire as it was deemed part of curriculum planning for senior leaders.
- Additional training and development questions for headteachers were removed as they were thought to overlap with previous answers. The question regarding strategies for managing and planning professional time (question 13; Annex 4: Final questionnaire) was retained, but only asked to headteachers as the issue was relevant at the school-level and not for individual teachers. This allowed a reduction in the overall survey length.

---

<sup>2</sup> See DfE microsite on the Workload Challenge: <https://www.gov.uk/government/publications/reducing-teachers-workload/reducing-teachers-workload>

- The test survey included the following questions about the use of resources.

*How often, if at all, do you use text books at [INSERT SCHOOL] for...?*

*Schemes of work; Individual lesson planning; Lesson delivery; Homework*

*How often, if at all, do you use teacher guides at [INSERT SCHOOL] for...?*

*How often, if at all, do you use online resources at [INSERT SCHOOL] for...?*

Questions about the use of text books were removed as teachers felt the questions could be construed as critical of teachers' professionalism and because of timing issues. As a result, this could have led to some teachers dropping out of the survey in response.

Similarly, the questions on use of teacher guides and online resources were removed, primarily for timing reasons, although they did also illicit similar discomfort in some cases.

- In completing grid questions on working environment (question 16, Annex 4: Final questionnaire<sup>3</sup>) and professional development (question 17, Annex 4: Final questionnaire<sup>4</sup>) it was felt that the survey was becoming onerous. A number of options were removed to reduce questionnaire length. These options were those which either had another item that could be used as a proxy, or that were thought to be less likely to contribute towards perceptions of workload as they were not cited in prior evidence as problematic.
- A question about the importance of addressing workload was removed:

*Taking all staffing issues in your school into consideration, how important, if at all, is it for you to address teachers' workload?*

This was deemed redundant since it was an obvious consequence of the previous question (seriousness of workload, question 19, Annex 4: Final questionnaire).

- The test survey included a satisfaction question with several aspects of teaching:

*How satisfied or dissatisfied are you with the following aspects of your role?*

*The number of hours I work;*

---

<sup>3</sup> Items removed:

Staff at [INSERT SCHOOL] have opportunities to contribute to strategic decisions affecting the school as a whole;

There is a mutually supportive environment between staff in the school;

Lesson observations carried out in [INSERT SCHOOL] are effective as part of professional development activity

<sup>4</sup> Items removed:

I have the skills I need to do my job effectively;

I have the tools I need to do my job effectively;

*The value society places in the teaching profession;*

*My own performance as a teacher / headteacher;*

*[TEACHER ONLY] My level of involvement in decisions that affect my work at [INSERT SCHOOL];*

*If I could decide again, I would still choose to work as a teacher.*

For several items, no direct relationship with workload could be made by teachers and so some items were removed as having limited value. Item D (my level of involvement in decisions that affect my work at [INSERT SCHOOL]) was moved to be included with question 18 of the final survey (Annex 4: Final questionnaire).

- Two advocacy questions about how respondents speak about being a teacher or about the teaching profession also proved quite emotive:

*On balance, which of the following statements comes closest to how you speak about being a teacher / about the teaching profession?*

*I speak positively about ... without being asked;*

*I speak positively about ... if I am asked about it;*

*I speak negatively about ... if I am asked about it;*

*I speak negatively about ... without being asked*

Teachers felt they could be perceived as unprofessional if they were negative. Teachers were passionate about their profession, but felt uncomfortable responding to how they would speak about it when asked. In the end, this question was removed in order to keep the questionnaire length down and to limit instances of drop out.

## **The Teacher Workload Survey 2016: Cognitive testing for schools**

### **About the Teacher Workload survey**

The Government wants to remove unnecessary workload in schools so that teachers can focus on improving outcomes for pupils. A range of actions have already been implemented or are in progress. Three review groups have been set up to address the causes of unnecessary workload in relation to marking, planning and resources and data management - the three biggest issues emerging from the Workload Challenge. The actions also include a commitment to monitor teacher workload through this biennial survey starting in spring 2016. [CFE Research](#) and the [Institute of Employment Relations at Warwick](#) are delivering the 2016 Teacher Workload Survey on behalf the Department for Education.

The Steering Group advising this study includes representatives from DfE, CFE Research, the University of Warwick and the following unions: ASCL, NAHT, ATL, NUT, The Voice Union, and NASUWT. Further information about the survey can be found at our website: <http://cfe.org.uk/our-work/teacher-workload-survey-2016>.

### **What is cognitive testing and what will it involve?**

Cognitive testing is an important part of this study and we are looking for schools to support this process. Cognitive testing involves a research team testing a survey that has been designed to ensure that it is working the way it should. If it doesn't it allows us to make changes before it is launched. For example it allows us to explore if questions are being understood by everyone in the same way and if question wording is clear. It is important to get this stage of the research right to make sure that questions are accurately measuring teacher workload.

If you agree to us undertaking cognitive testing at your school we would only need a small number of teachers to take part (approximately 3-4). They would be asked to complete the Teacher Workload survey online at the start of the session. This stage should take approximately 10 minutes. After this, we would then discuss the survey with them to get their thoughts on the survey completion process and the question wording so that we can be sure each question has been understood as we intended. We want to make sure that different teachers and different schools interpret the questions in the same way. We anticipate this discussion should take no more than 45 minutes, so we would need a maximum of an hour in total.

If you agree to take part we will visit your school or another venue of your choice. You will just need to ensure that each person taking part has separate internet access in order to take the survey (this can be on computers/laptops, tablets or mobile phones if preferred).

If you be happy to take part or would like more information on this please contact the Teacher Workload Survey team on the details below.

Teacher Workload Survey Team at CFE Research

[TeacherWorkload.Survey@cfe.org.uk](mailto:TeacherWorkload.Survey@cfe.org.uk)

Tel: 0116 229 3300

## Annex 4: Final questionnaire

1) In order to provide some context for the survey, can you say which of the following best describes your main role at [INSERT SCHOOL FROM SAMPLE]?

SINGLE CODE ONLY.

	Role	
A	Classroom Teacher	1
B	Head of Department	2
C	Head of Year	3
D	Deputy or Assistant Head	4
E	Headteacher / Acting Headteacher	5
F	Other (Please specify)	6

HIDDEN FIELDS

IF A TO C AND F = [TEACHER ROUTE](#)

IF D TO E = [HEADTEACHER ROUTE](#)

[FLAG SCHOOL PHASE FROM SAMPLE: SECONDARY OR PRIMARY](#)

\*\*\*\*\*

### Working hours questions – TEACHER route

[SHOW FOR TEACHERS](#)

The first questions cover the activities that are most common for classroom teachers [INSERT IF SECONDARY: , and heads of department and year].

Some questions will ask for an estimate of time spent in **hours** in your most recent full working week. “Full working week” means your last working week covering Monday to Sunday that was not shortened by illness, religious breaks or public holidays.

## ASK TEACHERS

- 2) In your most recent full working week, approximately how many hours did you spend in total on teaching, planning lessons, marking, covering for absence, interacting with other teachers, participating in staff meetings, pastoral care and other activities related to your job at [INSERT SCHOOL]?

Please include tasks that took place during weekends, evenings or other out-of-school hours.

Round to the nearest half hour. As an example, three and a half hours would be recorded as 3.5 below.

## RECORD HOURS

## ASK TEACHERS

- 3) Of this total, how many hours did you spend on teaching in your most recent full working week at [INSERT SCHOOL]?

Please only count actual teaching time.

Time spent on preparation, marking, etc. will be recorded later in the survey.

Round to the nearest half hour. As an example, three and a half hours would be recorded as 3.5 below.

Please record a 0 (zero) if you spent no time on teaching in your most recent full working week.

## ASK TEACHERS

4) How many hours did you spend on the following activities **other than teaching** in your most recent full working week at [INSERT SCHOOL]?

Include activities that took place during weekends, evenings or other off classroom hours.

Please exclude all time spent teaching.

Again, round to the nearest half hour. Please record a 0 (zero) if you spent no time on a listed activity.

RANDOMISE A TO J.

A	<b>Individual planning or preparation of lessons either at school or out of school</b>	Numeric	Hours
B	<b>Team work and dialogue with colleagues within this school</b>	Numeric	Hours
C	<b>Marking/correcting of pupils work</b>	Numeric	Hours
D	<b>Pupil counselling (including career guidance and virtual counselling)</b>	Numeric	Hours
E	<b>Pupil supervision and tuition (including lunch supervision)</b>	Numeric	Hours
F	<b>Pupil discipline including detentions</b>	Numeric	Hours
G	<b>Participation in school management</b>	Numeric	Hours
H	<b>General administrative work (including communication, paperwork, work emails and other clerical duties you undertake in your job as a teacher)</b>	Numeric	Hours
I	<b>Communication and co-operation with parents or guardians</b>	Numeric	Hours
J	<b>Engaging in extracurricular activities (e.g. sports and cultural activities after school)</b>	Numeric	Hours
K	<b>Other activities</b>	Numeric	Hours

## ASK TEACHERS

5) Across the whole school year, is the amount of time you spend on the activities outlined in the last question too little, too much or about right?

**SINGLE CODE ONLY FOR EACH ACTIVITY**

	<b>Statements</b>	<b>Far too little</b>	<b>Too little</b>	<b>About right</b>	<b>Too much</b>	<b>Far too much</b>	<b>n/a</b>
A	<b>Individual planning or preparation of lessons either at school or out of school</b>						
B	<b>Team work and dialogue with colleagues within this school</b>						
C	<b>Marking/correcting pupils' work</b>						
D	<b>Pupil counselling (including career guidance and virtual counselling)</b>						
E	<b>Pupil supervision and tuition (including lunch supervision)</b>						
F	<b>Pupil discipline including detentions</b>						
G	<b>Participation in school management</b>						
H	<b>General administrative work (including communication, paperwork, work emails and other clerical duties you undertake in your job as a teacher)</b>						
I	<b>Communication and co-operation with parents or guardians</b>						
J	<b>Engaging in extracurricular activities (e.g. sports and cultural activities after school)</b>						
K	<b>Other activities</b>						

## ASK TEACHERS

The next two questions ask more detailed activities relating to support, management and administration. When answering, it does not matter if you included any of the following in earlier responses you made.

6) How many hours did you spend on the following specific **support and management activities** in your most recent full working week at [INSERT SCHOOL]?

Round to the nearest half hour. Please record a 0 (zero) if you spent no time on a listed activity.

RANDOMISE A TO F.

A	<b>Non-regular teaching cover for absent colleagues within school's timetabled day</b>	Numeric	Hours
B	<b>Appraising, monitoring, coaching, mentoring and training other teaching staff</b>	Numeric	Hours
C	<b>Contact with people or organisations outside of school other than parents</b>	Numeric	Hours
D	<b>Organising resources and premises, setting up displays, setting up/tidying classrooms</b>	Numeric	Hours
E	<b>[ASK SECONDARY ONLY] Timetabled tutor time</b>	Numeric	Hours
F	<b>Staff meetings</b>	Numeric	Hours

## ASK TEACHERS

7) How many hours did you spend on the following specific **administrative activities** in your most recent full working week at [INSERT SCHOOL]?

Round to the nearest half hour. Please record a 0 (zero) if you spent no time on a listed activity.

RANDOMISE A TO C.

A	<b>School policy development and financial planning</b>	Numeric	Hours
B	<b>Recording, inputting, monitoring and analysing data in relation to pupil performance and for other purposes</b>	Numeric	Hours
C	<b>Planning, administering and reporting on pupil assessments</b>	Numeric	Hours

## ASK TEACHERS

8) **Across the whole school year**, is the amount of time you spend on the support and management activities outlined in the last questions too little, too much or about right?

SINGLE CODE ONLY FOR EACH ACTIVITY

	Statements	Far too little	Too little	About right	Too much	Far too much	n/a
A	<b>Non-regular teaching cover for absent colleagues within school's timetabled day</b>						
B	<b>Appraising, monitoring, coaching, mentoring and training other teaching staff</b>						
C	<b>Contact with people or organisations outside of school other than parents</b>						
D	<b>Organising resources and premises, setting up displays, setting up/tidying classrooms</b>						
E	<b>[ASK SECONDARY ONLY] Timetabled tutor time</b>						
F	<b>Staff meetings</b>						

ASK TEACHERS

9) **Across the whole school year**, is the amount of time you spend on the administrative activities outlined in the last questions too little, too much or about right?

	Statements	Far too little	Too little	About right	Too much	Far too much	n/a
A	School policy development and financial planning						
B	Recording, inputting, monitoring and analysing data in relation to pupil performance and for other purposes						
C	Planning, administering and reporting on pupil assessments						

\*\*\*\*\*

## Working hours questions – HEADTEACHER route

SHOW FOR HEADTEACHERS

The first questions cover the activities that are most common for Headteachers and Deputy Headteachers.

Some questions will ask for an estimate of time spent in **hours** in your most recent full working week.

“Full working week” means your last working week covering Monday to Sunday that was not shortened by illness, religious breaks or public holidays.

## ASK HEADTEACHERS

**10) In your most recent full working week, approximately how many hours did you spend in total on school management, staff supervision, interacting with other teachers, teaching and on other tasks related to your job at [INSERT SCHOOL]?**

Please include tasks that took place during weekends, evenings or other out-of-school hours.

Round to the nearest half hour. As an example, three and a half hours would be recorded as 3.5 below.

RECORD HOURS

## ASK HEADTEACHERS

### 11) How many hours did you spend on the following activities in your most recent full working week at [INSERT SCHOOL]?

Please include tasks that took place during weekends, evenings or other out-of-school hours. Again, round to the nearest half hour.

Please record a 0 (zero) if you spent no time on a listed activity.

DO NOT RANDOMISE.

A	<b>Leadership and management within the school</b> <i>Including strategic planning, staff meetings and other school-centred management activities, such as those associated with the management of federated schools.</i>	Numeric	Hours
B	<b>Administration within the school</b> <i>Including applying regulations to the school, reporting, school budget, preparing timetables and class composition.</i>	Numeric	Hours
C	<b>Administrative and management with external bodies</b> <i>Including responding to requests from local, regional, or national education officials</i>	Numeric	Hours
D	<b>Performance management of staff</b> <i>Including human resource/personnel issues, classroom observations, mentoring, initial teacher training and continuing professional development</i>	Numeric	Hours
E	<b>Teaching and related tasks</b> <i>Including covering for teacher absences, lesson planning, assessing and marking pupils' work and student assessment</i>	Numeric	Hours
F	<b>Curriculum planning</b> <i>Including developing curriculum and student evaluation</i>	Numeric	Hours
G	<b>Data analysis</b> <i>Including analysis performance data at the level of the teacher and the school and record keeping for external bodies / regulatory purposes</i>	Numeric	Hours
H	<b>Student interactions</b> <i>Including counselling and conversations outside structured learning activities, discipline</i>	Numeric	Hours
I	<b>Parent or guardian interactions</b> <i>Including formal and informal interactions</i>	Numeric	Hours
J	<b>Recruitment</b> <i>For teaching and support staff</i>	Numeric	Hours
K	<b>Other activities</b>	Numeric	Hours

ASK HEADTEACHERS

12) **Across the whole school year, is the amount of time you spend on the activities outlined in the last question too little, too much or about right?**

SINGLE CODE ONLY FOR EACH ACTIVITY

	<b>Statements</b>	<b>Far too little</b>	<b>Too little</b>	<b>About right</b>	<b>Too much</b>	<b>Far too much</b>	<b>n/a</b>
A	<b>Leadership and management within the school</b>						
B	<b>Administration within the school</b>						
C	<b>Administrative and management with external bodies</b>						
D	<b>Performance management of staff</b>						
E	<b>Teaching and related tasks</b>						
F	<b>Data analysis</b>						
G	<b>Curriculum planning</b>						
H	<b>Student interactions</b>						
I	<b>Parent or guardian interactions</b>						
J	<b>Recruitment</b>						

\*\*\*\*\*

## ASK HEADTEACHERS

13) Does [INSERT SCHOOL] have any of the following strategies in place for managing and planning professional time?

SINGLE CODE EACH. RANDOMISE A TO E

	Strategy	Yes	No	Not sure
A	Protected blocks on non-teaching time to plan lessons and/or mark work (PPA)			
B	Working collaboratively with other staff to plan schemes of work and/or share resources			
C	Existing schemes of work and associated lesson plans that can be adapted by teaching staff			
D	Computer software that effectively helps with administrative tasks			
E	A committee in place that monitors teachers' workload			
F	Other time management strategies (please specify)			

## Perceptions and drivers of workload

ASK ALL

14) To what extent do you agree or disagree with the following statements about your working hours?

SINGLE CODE EACH. RANDOMISE A TO C

	Statements	Strongly disagree	Tend to disagree	Neither agree nor disagree	Tend to agree	Strongly agree
A	I can complete my assigned workload during my contracted working hours					
B	I have an acceptable workload					
C	<u>Overall</u> , I achieve a good balance between my work life and my private life					

ASK ALL

15) You said earlier that you worked [INSERT RESPONSE FROM Q2 OR Q10] hours in your last working week. How many of those hours were spent working during weekends, evenings or other out-of-school hours?

Round to the nearest half hour. As an example, three and a half hours would be recorded as 3.5 below.

RECORD HOURS

ASK ALL

16) To what extent do you agree or disagree with the following statements about the working environment within [INSERT SCHOOL]?

SINGLE CODE EACH. RANDOMISE A TO D

	Statements	Strongly disagree	Tend to disagree	Neither agree nor disagree	Tend to agree	Strongly agree
A	Teaching staff collaborate effectively to address disciplinary problems					
B	Lesson observations carried out in [INSERT SCHOOL] are an effective part of professional development activity					
C	Teaching assistants are effectively deployed at [INSERT SCHOOL]					
D	Teaching staff collaborate effectively on teaching and learning					

ASK ALL

17) To what extent do you agree or disagree with the following statements about professional development and support?

SINGLE CODE EACH. RANDOMISE A TO E

	<b>Statements</b>	<b>Strongly disagree</b>	<b>Tend to disagree</b>	<b>Neither agree nor disagree</b>	<b>Tend to agree</b>	<b>Strongly agree</b>
A	<b>I have enough time to keep informed on changes to guidance and rules affecting professional practice</b>					
B	<b>I have the necessary Information and Communication Technology (ICT) skills to perform data recording and analysis tasks</b>					
C	<b>[INSERT SCHOOL] supports continuing professional development for teachers</b>					
D	<b>I have time during my contracted working hours to take part in professional development activities</b>					
E	<b>The resources available at my school to help plan teaching and learning are high quality</b>					

ASK ALL

18) To what extent do you agree or disagree with the following statements about the way you are managed?

SINGLE CODE EACH. RANDOMISE A TO C

	Statements	Strongly disagree	Tend to disagree	Neither agree nor disagree	Tend to agree	Strongly agree
A	[TEACHER ONLY] My manager is considerate of my life outside work					
B	[TEACHER ONLY] My manager recognises when I have done my job well					
C	I think that my performance is evaluated fairly					
D	I am satisfied with the number of hours I usually work					
E	[TEACHER ONLY] I am satisfied with my level of involvement in decisions that affect my work at [INSERT SCHOOL]					

ASK ALL

19) To what extent, if at all, do you consider teacher workload to be a serious problem in your school?

SINGLE CODE. REVERSE A TO D

A	Workload is a very serious problem	1
B	Workload is a fairly serious problem	2
C	Workload is not a very serious problem	3
D	Workload is not a serious problem at all	4
E	Don't know	5

SHOW FOR ALL

[i6] DfE would like to understand how workload, and other issues covered by this research, affects teachers' careers. To do this DfE would like its contractors to link teachers' answers to this survey to information held on the School Workforce Census, for the purposes of this research. This will provide further information about you and your career without you being asked further questions. Only aggregated statistics will be produced with this information, no individual or school will be identified in analysis or reports. Further information on how they will do this and how your privacy will be protected is available <here>.

**ASK ALL**

**20) Do you give permission for the research contractors to match your survey answers to your personal information held on DfE's School Workforce Census for this statistical analysis only?**

SINGLE CODE ONLY.

	<b>Response</b>	
A	<b>Yes</b>	1
B	<b>No</b>	2

**ASK IF YES AT Q20**

**21) In order for us to link to the School Workforce Census, can you please provide the following details:**

ALLOW REFUSED AS WELL AS DK FOR TEACHER REFERENCE NUMBER

	<b>Response</b>		<b>ALLOW REFUSED</b>
A	<b>Full Name</b>	OPEN TEXT FIELD	
B	<b>Date of birth</b>	USE DATE FIELD	
C	<b>Teacher Reference Number</b>	OPEN TEXT FIELD	

ASK IF NO AT Q20 OR REFUSED Q21

22) Into which of the following age bands do you fall?

SINGLE CODE ONLY.

	Age	
A	Under 25	1
B	25 to 29	2
C	30 to 34	3
D	35 to 39	4
E	40 to 44	5
F	45 to 49	6
G	50 to 54	7
H	55 to 59	8
I	60 or older	9
J	Prefer not to say	10

ASK ALL

23) What is your gender?

	Sex	
A	Female	1
B	Male	2
C	Prefer not to say	3

IF NO AT Q20 OR REFUSED AT Q21

24) Please provide the number of hours per week for which you are contracted to work at [INSERT SCHOOL]? If your contract is variable and/or term-time only, please write in the typical hours you work per school week.

RECORD HOURS

IF NO AT Q20 OR REFUSED AT Q21

25) Which of the following best describes your employment contract at [INSERT SCHOOL]?

SINGLE CODE ONLY.

A	<b>Permanent employment (an on-going contract with no fixed end-point before the age of retirement)</b>
B	<b>Fixed-term contract for a period of more than 1 school year</b>
C	<b>Fixed-term contract for a period of 1 school year or less</b>
D	<b>Don't know</b>

ASK ALL SECONDARY TEACHERS

26) Which of the following subjects do you teach on a regular basis?

MULTICODE. ALPHABETICAL ORDER DROP DOWN MENU

1	Applied Business Studies
2	Art and Design / Art
3	Biology / Botany / Zoology / Ecology
4	Chemistry
5	Commercial and Business Studies/Education/Management
6	Design and Technology
7	Design and Technology - Food Technology
8	Design and Technology - Resistant Materials
9	Drama
10	English
11	French
12	Geography
13	German
14	Health and Social Care
15	History
16	Humanities
17	Information and Communication Technology
18	Mathematics / Mathematical Development (Early Years)
19	Media Studies
20	Music
21	Other
22	Personal Social and Health Education (PSHE)
23	Physical Education / Sports
24	Physics
25	Primary Curriculum
26	Psychology
27	Religious Education
28	Science
29	Sociology
30	Spanish
31	Other (Please specify)
32	Prefer not to say

ASK ALL

27) **Are you a...?**

SINGLE CODE ITEM.

	<b>Role</b>	<b>Yes</b>	<b>No</b>
A	<b>... SEN Coordinator?</b>		
B	<b>... Newly qualified teacher (NQT)?</b>		

ASK ALL

28) **How many years you have been working in...?**

Please record the time in years to the nearest half year.

A	<b>...the teaching profession</b>	Numeric	Years
B	<b>...your current school</b>	Numeric	Years
C	<b>...your current role</b>	Numeric	Years

SHOW FOR THOSE WHO SAID YES AT Q20

DfE may wish to track the progress of your employment in teaching over the next three years (until 31st August 2019) to help better understand the any impact of workload over time. To do this DfE will keep a record of your survey answers and your School Workforce Census record. The data will only be used for research and statistical purposes.

Only aggregated statistics will be produced with this information, no individual or school will be identified in analysis or reports. Further information on how they will do this and how your privacy will be protected is available <here>.

29) **Do you provide permission for CFE Research to pass on a copy of your survey responses plus your personal details to DfE for this purpose?**

SINGLE CODE ONLY

	<b>Response</b>	
A	<b>Yes, you may pass on my linked survey and personal details to DfE to help them understand the longer term impact of workload on teachers' careers only</b>	1
B	<b>No</b>	2

SHOW TO ALL

DfE may also wish to contact you again in the next 12 months to ask you to take part in other research studies about teaching and workload issues. (Please note that, if

contacted, you will be under no obligation to take part). Further information on how your privacy will be protected is available <here>.

DfE may contact you directly or appoint an approved contractor. Your details would not be shared with any other third parties.

ASK ALL

30) Do you provide permission for CFE Research to pass on a copy of your survey responses plus your personal details to DfE for this purpose?

SINGLE CODE ONLY

	Response	
A	Yes, DfE may contact me to learn more about my survey responses and/or to invite me to take part in further research on this issue	1
B	No	2

SHOW IF YES AT Q20 AND YES AT Q30

31) Could you please provide details of how we can contact you for further research on teachers' workload. Note your contact details will not be used for any other purpose.

	Response		ALLOW REFUSED
A	Telephone number	OPEN TEXT FIELD	
B	Email address	OPEN TEXT FIELD	

SHOW IF NO AT Q20 AND YES AT Q30

32) Please provide some contact details so we can contact you for further research on teachers' workload. Note your contact details will not be used for any other purpose.

	Response		ALLOW REFUSED
A	Full Name	OPEN TEXT FIELD	
B	Date of birth	USE DATE FIELD	
C	Telephone number	OPEN TEXT FIELD	
D	Email address	OPEN TEXT FIELD	

END

Thank you for taking the time to complete this survey. The next steps are for the collected data to be analysed leading to a comprehensive published report from DfE

about English teachers' workload in the Autumn of 2016. This would not be possible without you kindly taking the time to provide your details and opinions.

## Annex 5: Response rates

### School cooperation rate

A target of 250 participating schools was set at the outset of the study. This target was challenging given the short timeframe for the study as a whole, including a short four week period for recruitment. Based on prior experience, it was estimated that a minimum sample of 750 schools would be required. The timeframe of the study meant that there would be little time to draw a reserve sample once the cooperation rate was known. As a result, the reserve was drawn at the same time as the main sample. An initial sample of 600 schools was invited to participate and after reviewing the cooperation rate, a further 250 were contacted early in the recruitment process. This left a reserve of 150 schools, all of whom were contacted by email at the end of the first week of fieldwork. The contact details of one school were incorrect, which left a total of 899 that were invited in some way to participate. As noted earlier (Annex 2: Sampling), school sampling by phase was conducted disproportionately to ensure that there were enough participating schools within the primary and secondary phases to address potential issues of a small number of sampling units in the secondary phase if representative sampling was used. The sample design was balanced in such a way as to provide enough individual clusters to minimise bias whilst still delivering the robust sub-samples of teachers for comparative analysis.

In total, 244 schools agreed to take part in the study and at least one response was received from 218 of these. Of the 899 invited, 128 schools (14%) refused to take part in the study, which left an eligible sample of 771. Table A5.1 shows the cooperation rates by school in total and between phases. The cooperation rate achieved for the full sample was 24%; the cooperation rate for the sample excluding refusals was 28%.

**Table A5.1: School cooperation rates**

	Primary		Secondary		Total	
	n	%	n	%	n	%
Total sampled schools	524	100	375	100	899	100
Refused	66	13	62	17	128	14
<i>Sample excluding refused</i>	458	87	313	84	771	86
Agreed to participate	138	26	98	26	236	26
Schools with participating teachers	125	24	93	25	218	24
Known contact (Refused + agreed)	204	39	160	43	364	41
No school agreement / contact	320	61	215	57	535	60
Co-operation rates						
Full sample		24		25		24
All excluding refused		27		30		28

## Teachers' response rate

As noted in Table A5.1, 218 schools had at least one teacher completing the survey. The teacher response rate was, therefore, calculated by comparing the number of responses received (in total and by sub-group) compared to the population of teachers present in these schools (based on aggregating the number of staff working at each participating school as recorded in the School Workforce Census at that point in time (March 2016)).

The response rate in each school ranged from 1% to 100%. Of the 218 participating schools, just 10 had a response rate of less than 10%. Table A5.2 shows the overall response rate was 33.9%.

**Table A5.2: Overall teacher response rates**

	Primary	Secondary	Total
Teachers present in participating schools	2,295	7,091	9,386
Completed interviews	900	2,141	3,186
Response rate	39.2%	30.2%	33.9%

## Improving cooperation and response rates in subsequent surveys

The survey has successfully improved on both the achieved sample size (1,004 diaries in 2013) and the response rate (15%) achieved in the last diary study (TNS-BMRB, 2013). As a result, the robustness and value of sub-group analysis has been enhanced. However, there are always opportunities to improve studies and two measures that could be taken to improve response rates in subsequent studies of this type are provided below.

1. **Provide more preparation time for school recruitment.** Out of necessity, the study was commissioned to a short timetable. Ideally, at least twice the time would have been allotted to the recruitment phase (i.e. at least nine weeks would be preferable). This would have allowed a similar number of schools to be recruited using a smaller total sample prior to the start of fieldwork. For the 2016 study, the recruitment and fieldwork period had to overlap to meet the pre-Easter deadline, which meant that some schools were still being recruited during the second week in fieldwork. Assuming that the three-week fieldwork window would remain in follow-up studies, it would be preferable to ensure all recruitment was completed before interviewing began. In this way, any recruitment that took place during the fieldwork period would be replacement activity to cover instances of drop-out by important selection strata (i.e. primary / secondary; size of school; location; etc.).

2. **Increased support activity to improve response and cooperation rates.** The 33.9% response rate was achieved through simplifying the methodology compared to past diary surveys and removing as much work as possible in survey administration from those coordinating within schools. There was little time or budget available to deliver the same level of technical and administrative support that was provided to schools in the TALIS survey, which also used mixed mode options (online and postal) to boost response rates and included minimum response rate targets as part of each participating country's contract (OECD, 2014). The level of support provided in TALIS was, therefore, necessary to fulfil contractual requirements. In addition, the sample administration was more complex and the questionnaire was much longer.

However, some further resource in future workload surveys would help minimise the number of participating schools with particularly low response rates (which would be excluded from analysis in TALIS) by providing more and targeted guidance and support if needed. For example, an experiment was run on a randomised sub-sample as part of the recruitment work. Half of the subset of schools were initially emailed by DfE and half by CFE Research. The DfE email was twice as effective in eliciting a positive response for cooperation from schools. Emails from DfE used as reminders were also fairly effective at increasing the number of surveys completed by teachers.

The contractor also had limited time to fully contact all of the sample due to the tight timetable. Whilst it was possible to call most schools at least three times within budget, it would have been useful to spend a little more time maximising coverage of the sample to increase the cooperation rate. Coupled with a more favourable timetable, there is no reason why a modest increase in the support budget could not improve both school cooperation rates and teacher response rates.

## Annex 6: Data processing and tabulation

### Cleaning

The use of Confinity survey scripting software ensured that necessary data cleaning was kept to a minimum – for example teachers were routed correctly through the survey depending on classroom teacher or head teacher role so that the right questions were answered. Closed questions were mostly compulsory until the demographic information section. In other regards data was kept as similar as possible to the TALIS survey for consistency and clarity of comparison.

Data cleaning was performed in the form of back-coding for the following questions:

- Q1 (role). In this particular instance, a small number of interviews were removed for those recording “Other (please specify)” because the response suggested that they were teaching assistants or other roles, which made them ineligible for the survey.
- Q4 / Q11 (hours spent on activities other than teaching). Hours spent on “other activities” were reassigned, where possible, to one of the ten pre-coded options..
- Q13 (other time management strategies). Open responses were back-coded as “Yes” into other pre-assigned categories where appropriate. All remaining responses were classed as “other”.

Additional data cleaning was performed where respondents had misinterpreted the question such as Q24 (number of contracted hours per week). Here, where people had given responses less than one (such as 0.8) it was reasonable to assume they had given their FTE equivalent. These were cleaned by multiplying their response by 32.5. This was done in order to derive a consistent figure in hours for as many respondents as possible. In the case of fractional responses to this question (0.1 to 0.9), it was deemed likely that teachers had given a response that reflected their hours proportional to full time hours. In cases where teachers gave permission for data-matching, it was possible to check data with the School Workforce Census and confirm this interpretation.

Finally, there were occasions during the “fuzzy matching” process of matching teachers to their School Workforce Census records, where it was clear that there were errors in date of birth given as there were discrepancies between dates entered into the survey and those held in the Census records. In these situations discretion was used to clean birth dates (for example where a year of birth was obviously given as “03” instead of “63”, or in cases where a given variable either did not match with the SWC record, or was inconsistent with other responses) so that age bands assigned to teachers were as accurate as possible. This process is described further below in the Data Matching section.

## Data Matching

Additional publicly available secondary data was included for all schools where available, specifically the Ofsted rating and region classification, pulled in from the Edubase database using the school's local authority establishment identifier (LAESTAB number).

In addition, teacher level data from the School Workforce Census (SWC) was included where the teacher gave consent at Q20/Q21. In this case it was found that the majority of consenting teachers were not aware of their Teacher Reference Number so they were matched with the secondary SWC data using a unique ID derived from teacher surname, date of birth and school identifier (LAESTAB).

This allowed matching of SWC data into the sample for the majority of those who gave consent. Where a match was not possible, "fuzzy matching" was utilised. This largely manual process involved identifying where close matches were present that could be confidently identified as correct matches, for example variations in spelling of surnames or easily identifiable errors in dates of birth.

The consent rate from teachers was less than expected; overall, 43% of respondents agreed, and 39% of respondents could be traced in the SWC and their data linked.

The main impact of the low match rate was that it placed a limitation on the extent to which it was possible to match teachers to obtain information for characteristics held on the SWC that were not collected as part of the survey. This was further complicated by the fact that there was a differential consent rate to matching by some teacher characteristics (i.e. younger age groups have a higher match rate than others). One potential action for future surveys is the introduction of an opt-out rather than an opt-in for data matching if such a question is permissible under data protection legislation.

## Processing and tabulating

The resulting cleaned data file had the weighting value appended (see separate documentation in Annex 7: Weighting for its calculation). In order to run cross-tabulations the data was converted from SPSS to Triple-S form and then onto an ASCII format readable by Merlin – a specialised data processing and tabulation tool widely used in Market Research.

A specification was documented identifying how each of the survey questions needed to be presented along with an exact description of the required cross-breaks. Once run, Merlin was then used to process and analyse the data according to the specification and this produced a complete cross-tabulation deck featuring all questions by the specified crossbreak.

Tabulations were run both unweighted and weighted and then thoroughly checked and validated against frequency counts and outputs produced from SPSS. This double-check ensured full confidence in data analysis produced via either method.

## Annex 7: Weighting

Weighting was necessary to account for unequal selection probabilities as part of the sampling process and to account for differential non-response from teachers based on a series of characteristics. Rim weighting<sup>5</sup> was completed on the achieved sample using six variables:

1. **Phase of School.** This made the largest contribution to the weights. The design of the study ensured enough secondary school clusters were present and, as a result, there were a disproportionately large number of secondary teachers in the sample compared to primary.
2. **Size of school.** The second variable that made a large contribution to the weighting was the size of the school. As above, the use of a census approach in the second stage of the PPS design (as opposed to selecting a fixed number of teachers per school) led to an over-representation of teachers from larger schools compared to the total population of all teachers.
3. **Role.** Binary classification of teacher (classroom, heads of department, heads of year) versus senior leaders (responsibility of assistant headteacher upwards). In the case of the secondary leader group, this was the largest difference in profile for all the survey weighting groups.
4. **Sex.** In the main, responses by sex were broadly representative.
5. **Age.** A binary weight was applied here: aged 34 or under; aged 35 plus. As per other DfE research, most newly qualified teachers (NQT's) sit within the first category (Adewoye, Porter and Donnelly, 2014).
6. **Contract.** Another binary classification between those on 0.8 FT contracts or above and 0.79 FT down. This fitted fairly well with the distribution found in the population profile.

Whilst data for all respondents existed for weighting variables 1 through 3 (as this formed part of the sample), not all 3,186 respondents gave an answer / could be matched to acquire sex, age and/or contract status. In these cases, a weight of 1 was assigned to that teacher.

The weighting was performed using SPSS and the syntax for this has been supplied to DfE as part of the agreed deliverables.

---

<sup>5</sup> Rim weighting is a process whereby an iterative process is used to weight survey data by more than one variable. It involves weighting data sequentially using a series of variables, then repeating the process in the same sequence until an equilibrium is reached whereby continued iterations no longer have an effect on the derived weights.

# Annex 8: Modelling the factors which impact on teacher workload

## Introduction to multilevel models

In Chapter 3, relationships between a range of explanatory variables and teacher workload were described. Many of these characteristics associated with workload are themselves interrelated. Further, an apparent relationship between a variable and workload may occur because both are related to a third variable, not considered in a simple two-way tabulation of percentages or mean values. For example, Goldstein (2003) gives examples where ignoring the nesting of pupil achievement scores into their particular schools results in estimated effects across pupils that don't accurately reflect the impact of school context. In the present analyses the relationships between self-reported total teacher workload and a range of explanatory variables were examined simultaneously. This minimised the possibility of being misled by spurious, inverse or absent associations which can occur when only one explanatory variable at a time is considered – which is an example of the Yule-Simpson paradox (Yule, 1903; Simpson, 1951, and discussed in the contemporary context of education research by Goltz & Smith, 2010).

## Between- and within-school variance components

The design of the present study involved the selection and testing of all teachers (where available) within each school. This induces an intra-cluster correlation (ICC) between teacher self-reported workload in schools. This means that the total variation in workload may be split into a between-cluster component, corresponding to school variation, and a within-cluster component, corresponding to individual teacher variation. The between-cluster variance component corresponds to the ICC expressed as a percentage. For classroom teachers the school-level variance component is 3.5% and for deputy/head teachers it is 15.4% (Table A8.1).

Postlethwaite (1995) suggests that large between-school variance components are suggestive of more heterogeneous school systems. The values reported here suggest a relatively homogeneous school population (in relation to teacher workload), with the majority of variation attributable to the teacher level. The school-level variance components were certainly much smaller than one would see for the variance components of pupil achievement, which would often be upwards of 33% (e.g., Cosgrove, J., Shiel, G., Sofroniou, N., Zastrutzki, S., & Shortt, 2005).

**Table A8.1: Variance components in total workload, by role**

<b>Role</b>	<b>Percent of Between-School Variance</b>	<b>Percent of Within-School Variance</b>	<b>Total</b>
Classroom	3.5	96.5	100.0
Deputy/head	15.4	84.6	100.0

Note: calculations are based on the final 2748 and 310 complete cases of classroom and deputy/head teachers, respectively, across the set of explanatory variables retained in the final models for each role.

## Procedure used for multilevel modelling

A form of multilevel model (Goldstein, 1987, 2003; Longford, 1993) known as a *hierarchical linear model* was used, as defined in Raudenbush and Bryk (2002). This is a type of regression model that contains a random component for each level. In its simplest form this will involve a normally distributed residual and a random intercept<sup>6</sup> for each nested level of clustering. In the 2016 Teacher Workload survey there were two levels – teacher’s self-reports, nested inside their respective schools. Estimated teacher effects may also be allowed to vary by fitting a random coefficient for a variable, which will suggest a range of likely values that the parameter estimate takes over the population of clusters (schools). Observations are assumed to be independent of one another once we condition on the random effects (i.e., the random intercept and any random coefficients) in the model. However, given the relatively small school-level variance components, the models are restricted to fixed coefficients using random intercept models which summarise the systematic trends in the data, while the variance components correctly adjust the standard errors of the estimates for the cluster sampled design<sup>7</sup>. The variables considered for evaluation in the model-building were those which were highlighted during the exploratory phase for chapter 2 showing associations with workload and/or which are of policy or theoretical interest, that are common to both classroom and deputy/head teachers. Where several variables were highly correlated or linked theoretically, the one with the strongest association with workload was chosen, or a composite variable considered. This reduces a variety of estimation issues amongst explanatory variables<sup>8</sup> and facilitates interpretation. The final set of candidate school- and

---

<sup>6</sup> The intercept varies across schools. This variation is normally distributed, and is summarised by a cluster-specific random effect, in addition to the normally distributed residual (within-cluster) variation that forms the kernel distribution of the model.

<sup>7</sup> The fixed coefficients also correspond to the mean of the corresponding school-level distribution of coefficients in a random coefficient model with multivariate Normal mixing distribution.

<sup>8</sup> Multicollinearity arises when explanatory variables are highly correlated, and leads to instability in the estimation of model terms (Hutcheson & Sofroniou, 1999).

teacher-level variables used in modelling is given in Table A8.2. These variables have been grouped using the levels shown and they were fitted by the R software using dummy variables with indicator coding.

**Table A8.2: Final set of teacher and school candidate variables**

<b>Candidate variables</b>
Teacher variable
Gender (female, male, prefer not to say)
Age in 2 categories (under 35, 35+).
NQT status or 3 terms of induction proxied by 1 year or less working in teaching (yes, no)
Number of years working in teaching (0-5, 6-10, 11+)
Full- or part-time status (full, part)
Teaching staff collaborate effectively to address disciplinary problems (5-point agreement scale)
Management: I think that my performance is evaluated fairly (5-point agreement scale)
School variable
Total number of teachers headcount in 2 categories (1-25, 26+)
School phase (primary, secondary)

Note: NQT status does not apply to deputy/head teachers and so it is not considered in the models for teachers in that role.

Following the advice in Aitkin, Francis, Hinde, and Darnell (2009), models were unweighted, as the explicit stratifying variables (school phase and size) were evaluated as part of model development. This is known as a disaggregated approach and contrasts with the weighted approach used to estimate a summary quantity when aggregating over the design variables, as in the descriptive statistics presented in chapter 2.

Metric variables which are continuous or with a large numbers of levels have been grouped into a smaller number of ordered-levels. This allows convenient tabulations as well as direct estimates of their association with workload by comparisons between the different levels, e.g., number of years working in teaching has three levels (0-5, 6-10, 11+). To allow straightforward comparisons, categorical explanatory variables (known as factors) have their reference levels set to a convenient group for comparisons among levels. The grouped numerical variables were treated in the same fashion, e.g., number of years working in teaching was represented explicitly by the two contrasts 0-5 versus 11+ years and 6-10 versus 11+ years. However, other comparisons are also possible by first calculating the fitted values for two groups of teachers (explained below in the section entitled ‘Contributions of the Explanatory Variables to Fitted Scores’) and then subtracting one from the other to provide the estimated difference in workload for the two groups being compared.

Models were compared by means of the *Akaike Information Criterion* (AIC), which adds a penalty of twice the number of parameters in the model to the *deviance* statistic<sup>9</sup> (Akaike, 1973). This is a measure specifically designed for the repeated process of testing terms during model-building, where good-fitting models are emphasised so long as not too many parameters are fitted.

Generally one would prefer a model with the smallest value of AIC, though differences within 1-2 points of the 'best' deserve consideration and those with differences of 3-7 points have considerably less support (Burnham & Anderson, 1998). An account of AIC and its advantages over traditional significance tests in the context of model-building is given in Lindsey (2004). One key advantage is the straightforward comparison of models that are not nested one within the other. In addition, two-way interactions at the teacher-level and cross-level interactions between cluster-level (school) variables and teacher-level variables were evaluated using the AIC. Actual AIC values are reported, rather than differences compared to some reference model, since comparison is possible amongst various models, whether nested within each other or not.

Separate random-intercept models of achievement for each explanatory variable fitted alone were initially constructed to evaluate whether, when variables were later entered simultaneously, the parameter estimates had changed substantially. Such change would indicate that the explanatory variables in the model were related in a complex manner and that the parameter estimates were sensitive to the other variables present. Full maximum likelihood estimation was implemented as it allows the AIC to be used to evaluate fixed effects as well as any random effects considered for the model. Categorical variables were examined by means of omnibus or overall changes in AIC, in which the model was fitted both with and without the corresponding set of dummy variables. The NLME library of Pinheiro and Bates (2000) implemented in the R statistical package was used to fit the multilevel models.

Each two-way interaction (teacher-level and cross-level) was examined, one at a time, by addition to the final main-effects model. Any significant interactions were then evaluated by simultaneous addition to the main effects model. Interactions explored were limited in advance to those terms involving school phase with the final variables, which kept the task well focused on a key system variable, helping to minimize any tendency towards "data dredging". Inset A8.1 provides some guidance for the interpretation of the more complex information presented in this chapter.

#### ***Inset A8.1. Interpreting the Tables of Multilevel Models***

The following points may be borne in mind when interpreting the tables in this chapter.

---

<sup>9</sup> The deviance is a measure of the probability of obtaining the observed data, given the fitted model, i.e.,  $-2 \log$  likelihood.

- The estimates in all tables are unweighted. When variables are added to the null (empty) model separately (e.g., Table A8.3), because the estimates are unweighted, they do not correspond exactly to the estimates provided in Chapter 2. The listwise deletion of cases with missing values (i.e., removal of all cases with one or more missing values on the explanatory variables) implemented in the *R* software will also lead to a difference in the estimates.
- Continuous and metric variables, such as age, have been grouped into a smaller number of levels, which provides a straightforward comparison with the effects estimated for categorical variables.
- For categorical variables (such as school phase), the chosen reference category is given alongside the label for the category corresponding to the parameter estimate.
- In describing the tables, the parameter estimates correspond to contrasts between the indicated level of the variable and the reference category.
- The parameter estimate for the intercept corresponds to the mean workload for a teacher with the range of characteristics denoted by the reference category of each variable.
- Fitted values for teachers with other characteristics are obtained by adding the parameter estimate for that attribute to the intercept.
- When interpreting the variance components for the final models (such as that for classroom teachers shown in Table A8.1), it is useful to note that the square root of the variance is the standard deviation. For example, the variance component for the intercept of the final model of classroom teachers is 2.92; the standard deviation is 1.71.
- With random intercepts, the standard deviation can be used to calculate the proportion of schools expected within a given range of mean scores. As the models utilised fit Normal distributions to explain the variation in schools, 68% of schools can be expected to fall within a range of + or – 1 standard deviations of the particular fitted value corresponding to any given set of teacher and school characteristics.

## Explanatory variables evaluated as separate models

Tables A8.3 and A8.5 present the coefficients for each teacher-level variable evaluated as separate models for the classroom and deputy/head teachers, respectively. The corresponding coefficients covering the school-level variables are given in tables A8.4 and A8.6. For comparison, the AIC of the null random-intercept models is given. As the datasets are different the AIC values should only be compared among models fitting workload for a given teacher role, e.g., comparing the fit of different models within those addressing classroom teachers. The variable “teaching staff collaborate effectively to address disciplinary problems” showed little

evidence of an association with teacher workload, for classroom teachers or deputy/head teachers and it is not discussed further.

A model for classroom teachers with teacher headcount alone (AIC=20082.8) does not improve the penalized model fit over the null random-intercept model (AIC=20081.2). This can also be seen in the standard error for teacher headcount, which exceeds its parameter estimate<sup>10</sup>. In contrast, models such as those with Gender, Age, NQT status and full- or part-time status fitted as separate models, show improvements in the model fits that are worthwhile. All variables used in the modelling are included at the foot of each explanatory table in the following section (Tables A8.3 through A8.8).

---

<sup>10</sup> A common guideline is that a parameter estimate/standard error with an absolute value exceeding 2 provides considerable support for the retention of a variable in the model. This corresponds to a t-test with moderate-to-large degrees of freedom. For the summary material presented in chapter 6, the numerical values of parameter estimates were suppressed and replaced with the text “*no significant difference*” when the t-ratios were considerably less than 2 in value, e.g., Gender: *prefer not to say* versus *female*.

**Table A8.3: Classroom teachers' total workload: Teacher-level variables tested as separate models by addition to the null random intercept model**

	<b>Parameter</b>	<b>Standard Error</b>	<b>AIC</b>
Intercept only	53.954	0.303	20081.2
Gender			20058.2
male–female	2.769	0.532	
prefer not to say–female	0.503	1.291	
Age, 35+–under 35	-2.176	0.476	20062.4
NQT status, no–yes	-2.893	0.884	20072.5
Number of years working in teaching			20030.5
0 to 5–11+	4.097	0.556	
6 to 10–11+	1.136	0.583	
Full- or part-time status, part–full	-14.275	0.567	19516.0
My performance is evaluated fairly			20072.5
strongly disagree–neutral	2.684	1.118	
disagree–neutral	1.946	0.791	
agree–neutral	0.009	0.633	
Strongly agree–neutral	1.781	0.850	

Note: “Teaching staff collaborate effectively to address disciplinary problems” showed little evidence of an association with teacher workload and is omitted from the table.

**Table A8.4: Classroom teachers' total workload: School-level variables tested as separate models by addition to null random intercept model**

	<b>Parameter</b>	<b>Standard Error</b>	<b>AIC</b>
Intercept only	53.954	0.303	20081.2
School phase, primary–secondary	1.191	0.641	20079.8
Teacher headcount, 26+–1 to 25	0.434	0.697	20082.8

In contrast to the data for classroom teachers, there are fewer of the candidate variables which when fitted separately as models of deputy/head teachers workload improve upon the null random intercept model. A model for deputy/head teachers with Number of years working in teaching by itself (AIC=2001.7) does not improve the penalized model fit over that obtained for the null random-intercept model (AIC=1999.8). This can be confirmed by considering the standard error for Number of years working in teaching, level 6 to 10 versus 11+, which exceeds the

absolute value of its parameter estimate<sup>11</sup>. In such cases the estimated parameters represent arbitrary “noise” rather than an informative “signal” and indicate that the variable is not a good predictor of workload. The model with Full- or part-time status clearly improves the model fit considerably over the null random intercept model. In addition, the magnitude of the parameter estimates compared to the corresponding standard errors for Gender, level male versus female, and School phase, primary versus secondary, suggest further exploration in a fuller model. As it is possible for variables to play a more complex role, when fitted in combination, exploration was continued in the context of models involving multiple variables.

**Table A8.5: Deputy/head teachers’ total workload: Teacher-level variables tested as separate models by addition to null random intercept model**

	<b>Parameter</b>	<b>Standard Error</b>	<b>AIC</b>
Intercept only	60.736	0.667	1999.8
Gender			2000.6
male–female	2.305	1.304	
prefer not to say–female	3.158	5.862	
Age, 35+–under 35	1.014	2.046	2001.6
Number of years working in teaching			2001.7
0 to 5–11+	N/A	N/A	
6 to 10–11+	-0.799	1.916	
Full- or part-time status, part–full	-7.649	2.696	1993.9
My performance is evaluated fairly			2007.0
strongly disagree–neutral	3.6487	5.230	
disagree–neutral	1.883	3.348	
agree–neutral	2.316	2.729	
Strongly agree–neutral	2.275	2.741	

Note: “Teaching staff collaborate effectively to address disciplinary problems” showed little evidence of an association with teacher workload and is omitted from the table. NQT status is not applicable for this role and has not been considered, here.

**Table A8.6: Deputy/head teachers’ total workload: School-level variables tested as separate models by addition to null random intercept model**

	<b>Parameter</b>	<b>Standard Error</b>	<b>AIC</b>
Intercept only	60.736	0.670	1999.8
School phase, primary–secondary	-1.513	1.338	2000.6
Teacher headcount, 26+–1 to 25	0.655	1.354	2001.6

<sup>11</sup> There is no parameter estimated for 0 to 5–11+ as the younger level, 0 to 5, is not applicable to deputy/head teachers in the dataset.

## The final models of total workload

All the candidate teacher-level variables were fitted simultaneously in initial model fits: that is, for classroom teachers, Gender, Age, NQT status, Number of years working in teaching, Full- or part-time status and the response to the management item My performance is evaluated fairly.

To these variables, the school-level variables School phase and Teacher headcount were added.

The model fitted considerably better than the any of the separate models (AIC=19482.0).

However, omitting each variable in turn and then replacing it suggested the following sequence of elimination of model terms: Age and NQT status. This resulted in a reduced model containing the following terms: Gender, Number of years working in teaching, Full- or part-time status, My performance is evaluated fairly, School phase and Teacher headcount. Each two-way interaction with School phase, e.g., Gender x School\_phase, was then added to the model, one at a time but no improvement in the penalized model fit indicated by the AIC values was suggested. The AIC of this final main-effects random intercept model for classroom teachers was 19480.1.

Subsequently the model was refitted to the larger subset of teachers who gave complete responses to the final six variables and the parameter estimates and standard errors are given in Table A8.7.

**Table A8.7: Final model of classroom teachers' total workload**

	<b>Parameter</b>	<b>Standard Error</b>	<b>AIC</b>
			19480.1*
Intercept	52.093	1.0128	
Gender			
male–female	1.434	0.485	
prefer not to say–female	0.906	0.984	
Number of years working in teaching			
0 to 5–11+	1.933	0.503	
6 to 10–11+	0.230	0.520	
Full- or part-time status, part–full	-13.799	0.571	
My performance is evaluated fairly			
strongly disagree–neutral	1.892	0.982	
disagree–neutral	1.629	0.697	
agree–neutral	-0.108	0.558	
Strongly agree–neutral	0.597	0.752	
School phase, primary–secondary	4.032	0.825	
Teacher headcount, 26+–1 to 25	2.107	0.880	
Variance components			
Intercept	2.923		
Residual	110.195		
Variables dropped from the model (in sequence):			
Age, NQT status			

Note: the parameter estimates and standard errors are for the subset of 2622 complete cases for these variables. \*The AIC value is for the same model, but fitted to the complete cases for the larger initial variable set to enable comparison with the simpler models.

For deputy/head teachers the initial complete main effects model which was explored included Gender, Age, Number of years working in teaching, Full- or part-time status and the response to the item My performance is evaluated fairly. The school-level variables School phase and Teacher headcount were also included. This initial model yielded a penalized model fit that was no better than several of the separate main effects models (AIC= 2008.5), indicating that it included an excess of model terms and was over fitting the data. Omitting each variable in turn and then replacing it suggested the following sequence of elimination of model terms: My performance is evaluated fairly, Number of years working in teaching and Teacher headcount. This resulted in a considerably reduced model containing the following terms: Gender, Age, Full- or part-time status and School phase. The two-way interactions with School phase, e.g., Age x School\_phase, were explored as additions to the model, one at a time but no improvement was indicated. As the presence of School phase as a main effect increased the AIC value, it was omitted to yield the final model. The AIC of this final main-effects random intercept model for deputy/head teachers was 1996.5. It should be noted that a considerably simpler model with just Full- or part-time status is also supported by its AIC value, but given the smaller sample size of

deputy/head teachers (269 complete cases compared to 2622 classroom teachers available for the final model) and the exploratory nature of this analysis, the more complex model was preferred which is also given some support from the magnitude of the final variables' parameter estimates in relation to their standard errors. The final model with Gender, Age, Full- or part-time status was then refitted to the larger subset of teachers with complete responses to these three variables. The corresponding parameter estimates and standard errors are given in Table A8.8.

**Table A8.8: Final model of deputy/head teachers' total workload**

	<b>Parameter</b>	<b>Standard Error</b>	<b>AIC</b>
			1996.5*
Intercept	57.857	1.972	
Gender			
male–female	2.551	1.340	
prefer not to say–female	2.818	6.009	
Full- or part-time status, part–full	-7.242	2.794	
Age, 35+–under 35	2.458	2.041	
Variance components			
Intercept	15.771		
Residual	91.341		
Variables dropped from the model (in sequence):			
My performance is evaluated fairly, Teacher headcount, Number of years working in teaching, School phase			

Note: the parameter estimates and standard errors are for the subset of 269 complete cases for these variables. \*The AIC value is for the same model, but fitted to the complete cases for the larger initial variable set to enable comparison with the simpler models.

The formula for explained variance (based on Snijders & Bosker, 1999) is given in Inset A8.2 and uses a value for the typical group size. An estimate of the typical number of FTE teachers with that role in each school, i.e., 15 classroom teachers and 3 deputy/head teachers, was entered for this value accordingly. The final model for classroom teachers explains 29.0% of the between-school variation and 22.1% of the within-school variation (which is the much larger of the two variance components, see Table A8.1). In contrast, the much simpler final model of deputy/head teacher workload explains 5.6% of the between-school variation and 4.6% of the within-school variation. To some extent these latter values reflect the simplicity of the model, but the inherently small size of the deputy/head teacher clusters set at typically 3 per school does play a role here. The latter value is a reflection of the population size of the clusters for that sub-population rather than being a sample size limitation. Further, the total number of observations for the analysis of deputy/head teachers was limited to only 269 complete cases. This suggests that, with such small clusters and limited sample size, the multilevel estimate of their explained variances should be viewed with caution.

**Inset A8.2. Calculation of the Proportion of Explained Variance in Teacher Workload**

The method used to calculate the proportion of variance in achievement at level 2 (school level) requires one to use a representative value for the size of the level 2 clusters. An estimate of number of classroom teachers in a school – 15 FTE teachers – was used for the representative level 2 cluster size in the models of workload for that role. For deputy/head teachers the value was 3 FTE teachers. The formulae used were:

$$\text{Level 1 } R^2 = 1 - (\text{Var } L1F + \text{Var}L2F)/(\text{Var}L1N + \text{Var}L2N)$$

$$\text{Level 2 } R^2 = 1 - (\text{Var } L1F/CS + \text{Var}L2F)/(\text{Var}L1N/CS + \text{Var}L2N)$$

Where VarL1F = Level 1 variance of fitted model; VarL2F = Level 2 variance of fitted model;

VarL1N = Level 1 null model variance; VarL2N = Level 2 null model variance; CS = Cluster Size

For example, the calculation for the final model of classroom teacher workload uses VarL1F = 110.195, VarL2F = 2.923, VarL1N = 139.998, VarL2N = 5.121 and CS = 15.

Therefore

$$\text{Level 1 } R^2 = 1 - ((110.195 + 2.923) / (139.998 + 5.121)) = 0.221$$

and

$$\text{Level 2 } R^2 = 1 - ((110.195/15 + 2.923) / (139.998/15 + 5.121)) = 0.290$$

Further details are given in Snijders and Bosker (2012).

## Contributions of the explanatory variables to fitted scores

All the explanatory variables fitted with the present models consist of a small number of groups, either from splitting metric variables, e.g., age into levels corresponding to age bands, or inherently categorical ones such as gender. The parameter estimates consist of contrasts between a given level and a reference category chosen to facilitate interpretation. Due to the choice of reference category for each variable, the intercept for the classroom teacher model denotes the mean workload for a teacher with the following characteristics:

- Female;
- Number of years working in teaching: 11+;
- Full-time employment status;
- “My performance is evaluated fairly”: neutral response;
- School phase: secondary;
- Teacher headcount: 1 to 25.

From Table A8.7 it can be seen that this reference mean workload has a fitted value of 52.1 hours. The fitted values for other types of classroom teacher are obtained by simply adding the parameters estimates for the variables on which they differ to the intercept. For example, a classroom teacher who was Part-time in a Primary school would have a fitted workload value of  $52.1 + 4.0 - 13.8 = 42.3$  hours. Other combinations can be worked out similarly, but the tables can also be interpreted directly to consider the estimated effect of a change in the level of a variable compared to the reference category. For example, the estimate for being in a school with a Teacher headcount, 26+ has a fitted workload that is 2.1 hours more than the reference category 1 to 25. The net Gender difference is 1.4 additional hours reported by males and 0 to 5 years working in teaching has a difference of 1.9 hours compared to 11+ years. The management variable “My performance is evaluated fairly” suggested that teachers who disagree or strongly disagree with the statement report a workload that is 1.6 and 1.9 hours more than those who are neutral on the item, respectively, (with agree and strongly disagree not being significantly different from neutral). The self-reported additional workload for classroom teachers in Primary schools remains high at 4.0 hours after the effect of the other variables is taken into account by the statistical model.

**Table A8.9: Cross-classified percentages for full- or part-time status versus gender and age: obtained sample characteristics for deputy/head teachers.**

Age	Gender	Full- or part-time status			count
		full	part	total	
under 35	female	90.9	9.1	100.0	22
	male	100.0	0.0	100.0	7
	prefer not to say	...	...	...	0
35+	female	94.5	5.5	100.0	146
	male	94.5	5.5	100.0	91
	prefer not to say	100	0.0	100.0	3

Note: Cell values are based on unweighted data and correspond to the obtained sample.

In the case of the final model for deputy/head teachers the variables present and the chosen reference category result in the parameter estimate for intercept corresponding to the self-reported total workload for a teacher who is:

- Female;
- Under 35 years-old,
- Full-time employment status,

with a fitted value of 57.9 hours. Males report 2.6 hours more than females, older 35+ year-old workers are estimated to report 2.5 hours more than their younger counterparts, while part-time status corresponds to a fitted value that is 7.2 hours less than for full-time teachers.

The simpler final model for the deputy/head teachers is amenable to further exploration by an examination of the trivariate characteristics of the sampled teachers. As the explanatory variables are themselves interrelated we reproduce the relative frequencies for each in table A8.9 illustrating the changes in Gender composition for Full- or part-time status and across Age. In the obtained sample, male and female deputy/head teachers work full-time to similar high levels of around 94.5% in full-time employment for the older 35+ age group, but 90.1% of females work full-time compared to 100% of males in the younger under 35 age group. The patterns are derived from a base-size of 267 cases and should be considered indicative only, but they do shed some light on the characteristics of the teachers that underlie the estimates of workload differences from the final model. However, a net gender effect remains in the model after having included terms for Age and Full- or part-time status. One might hypothesize that female deputy/head teachers differ from males both in the nature of the post held, as well as perhaps in the efficiency with which they utilize their available time, given higher levels of additional unpaid non-teaching responsibilities at home. However, these are matters for further research.

The present analyses have compared the explanatory utility of a range of teacher and school factors in relation to their impact on self-reported total workload by classroom and deputy/head teachers considered separately. The variable set explored includes those factors which were common to both roles of teachers, enabling a comparative discussion of their effects across the models for the two groups. The small size of the school variance component, especially for classroom teachers, at 3.5% to the total variation in workload, points to some inherent limitations for policy interventions. Essentially, any policy that seeks to isolate a minority of high average-workload schools, e.g., those schools above the 95<sup>th</sup> percentile, will be greatly restricted in its potential impact as the vast majority of variance occurs at the individual teacher level rather than in particular schools. Effective interventions would need to target teachers across the population of schools.

More detailed secondary analysis of the Teacher Workload Survey dataset by future researchers could usefully explore the efficacy of role-specific explanatory variables in improving the fit of the models further, though the limits imposed by the considerably smaller sample size of the deputy/head teachers should be borne in mind in analyses of workload for that role.

## References

- Adewoye, M., Porter, S. and Donnelly, L (2014) Newly Qualified Teachers: Annual Survey 2014. DfE. London
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on inference theory* (p. 267-281). Budapest, Hungary: Akademiai Kiado.
- Aitkin, M., Francis, B., Hinde, J., & Darnell, R. (2009). *Statistical modelling in R*. Oxford, UK: Oxford University Press.
- Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: A practical information-theoretic approach*. New York: Springer.
- Cosgrove, J., Shiel, G., Sofroniou, N., Zastrutzki, S., & Shortt, F. (2005). *Education for life: The achievements of 15-year-olds in Ireland in the second cycle of PISA*. Dublin, Ireland: Educational Research Centre.
- Longford, N. T. (1993). *Random coefficient models*. Oxford, UK: Oxford University Press.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Charles Griffin & Company.
- Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.
- Goltz, H. Honoré., & Smith, M. L. (2010). Yule-Simpson's Paradox in Research. *Practical Assessment, Research & Evaluation*, 15(15). Available online:
- Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: Sage Publications.
- Krosnick, J. A., (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*. Volume 5, Issue 3. May/June 1991. Pages 213–236
- Micklewright, *et al* (2014) Teachers in England's Secondary Schools: Evidence from TALIS 2013, DfE, London. p.15
- OECD (2014) *TALIS 2013 Technical Report*. Table 5.3, p.80
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York: Springer.
- Postlethwaite, T. N. (1995). Calculation and interpretation of between-school and within-school variance in achievement. In OECD (Organisation for Economic Co-operation and Development) (Ed.), *Measuring what students learn* (pp. 81-91). Paris: Editor.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, 13 (2), 238-241.

Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modelling* (2nd ed.). London: Sage.

TNS-BMRB (2014) Teachers' workload diary survey 2013. Research Report. DfE. London. p.10.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, 2(2), 121-134.



Department  
for Education

© DfE

**Reference: DFE-RR632**

**ISBN: 978-1-78105-690-5**

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

Any enquiries regarding this publication should be sent to us at:

[TeachersAnalysisUnit.MAILBOX@education.gsi.gov.uk](mailto:TeachersAnalysisUnit.MAILBOX@education.gsi.gov.uk) or  
[www.education.gov.uk/contactus](http://www.education.gov.uk/contactus)

This document is available for download at [www.gov.uk/government/publications](http://www.gov.uk/government/publications)