Teacher workload and well-being. New international evidence from the OECD TALIS study.

John Jerrim
Sam Sims

UCL Institute of Education

November 2020

In many Western education systems, teachers are expected to work long hours during term-time. Yet little is currently known about what the implications of these working patterns are for teachers' well-being and mental health. This paper provides new evidence on this issue for five predominately English-speaking education systems (England, Australia, Alberta-Canada, New Zealand, United States). It illustrates how, by ignoring potential non-linearities and measurement error, previous research may have missed key aspects of the workload-wellbeing relationship. We also demonstrate how the time that teachers spend marking is the key driver of workload stress and poor levels of workplace wellbeing across English-speaking countries.

Key Words: Teachers, workload, stress, mental health, working hours, TALIS.

Contact details: John Jerrim (J.Jerrim@ucl.ac.uk) Department of Social Science, UCL Institute of Education, University College London, 20 Bedford Way London, WC1H 0AL

Acknowledgements: This research has been funded by the Nuffield Foundation. The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views of the authors and not necessarily the Foundation. Visit www.nuffiedlfoundation.org. Helpful comments have been received on the draft from our project steering group, who we would like to thank for their input and support.

1. Introduction

Recent research found that a quarter of teachers in England work more than 59 hours per week (Allen et al., 2019). These long hours have been blamed for high levels of work-related stress (National Education Union, 2018) and declining levels of retention among teachers (Foster, 2019). Headteachers report that this is hindering the quality of instruction provided by schools (Jerrim & Sims, 2019). Finding out more about the relationship between teacher workload and well-being is hence an issue of great education policy interest, while also being a topic of much debate amongst teachers and teaching unions. The aim of this paper is to provide new international evidence on this matter, tackling a number of empirical challenges that have not been considered within the existing literature.

There is a small but growing qualitative literature on working hours and the changing composition of workload within the teaching profession. This research tends to find that teachers are dissatisfied with their workload (Cooper-Gibson, 2018; Lam & Yan, 2011; Perryman & Calvert, 2019) but also emphasises that certain aspects of workload are viewed more negatively than others. In particular, the growing demands of assessment, marking and data entry, often in order to comply with (perceived) demands of accountability systems are particularly unpopular with teachers (Bradbury & Roberts-Holmes, 2018; Perryman & Calvert, 2019; Selwin, Nemorin, & Johnson, 2017).

While this literature provides useful insights on the relationship between workload and wellbeing, it also has some important limitations. First, the qualitative nature of the data limits the extent to which the relative importance of different components of workload can be established. For example, what is the effect on wellbeing of an additional hour spent upon marking and administration, relative to additional hour spent on contact time with pupils? Second, this research is largely focused on England. While some elements of teacher workload may be stressful in general, other parts of the job may only be linked with work related stress within certain national settings. Take, for example, the time teachers spend liaising with parents. In a country where teachers and parents have a collaborative and respectful relationship, this may be quite an enjoyable task. Yet in another country, where teachers only meet with parents to discuss challenging pupil behaviour, then time spent upon this activity might be rather stressful.

A separate strand of the literature, often drawing on concepts and methods from educational psychology, uses teacher survey data to investigate the relationship between workload and

wellbeing. This literature tends to find that teachers who experience greater demands in the workplace are more likely to experience burnout (Fernet et al., 2012; Hakanen, Bakker, & Schaufeli, 2006; Kokkinos, 2007) and reduced job satisfaction (Kinman, Wray, & Strange, 2011; Skaalvik & Skaalvik, 2009). Such research often stresses the importance of teachers acting autonomously - in the sense of endorsing the value of the tasks they are required to undertake – rather than doing them solely in order to comply with outside demands (Fernet et al., 2011; Skaalvik & Skaalvik, 2009). This further reinforces the value of separately investigating the relationship between different aspects of workload and teacher wellbeing.

While valuable, this strand of the literature also has its limitations. Foremost is that the questionnaire items measuring workload tend to rely on teachers' self-reported perceptions of whether their workload is overly demanding. The absence of measures of the amount of time spent on different tasks limits the type of inferences that can drawn about what constitutes *too much* workload. In addition, this type of research tends to utilise small convenience samples, which calls into question the representativeness of the findings.

The present research aims to address some of these gaps in the literature by drawing on the 2018 edition of the Teaching and Learning International Survey (TALIS). In particular, we analyse representative data on more than 10,000 teachers from five English-speaking jurisdictions: Australia, England, New Zealand, United States and Alberta-Canada. This allows us to investigate how the time teachers spend upon a selection of different tasks is related to self-reported work-related stress and workplace well-being, across a range of national settings.

In addition to utilising this newly released data source, we make two methodological advances. First, existing research implicitly assumes a linear relationship between workload and wellbeing: that each additional hour spent working (or each additional hour spent upon a particular task) has the same marginal impact upon teacher well-being. Yet there is little reason to believe that this is necessarily the case. Spending one more hour marking per week may only be a minor nuisance if your existing workload is low, but could cause you great stress if you already have a very high workload. We address this by running variations of our main models using non-parametric regression. Second, the existing literature tends to gloss over the empirical challenges in estimating the association between wellbeing and workload when the latter is notoriously difficult to measure (Allen et al., 2019). We address this by using structural equation modelling to investigate how our estimates vary under different assumptions about the reliability with which workload is measured.

To trail our key findings, we show that longer working hours are associated with higher levels of workload stress, and lower levels of workplace wellbeing, across all the countries considered. Indeed, our analysis suggests that much of the previous literature on this issue, which has implicitly assumed that teachers' self-reported working hours are measured without error, may have underestimated the strength of this relationship. There is also evidence that, in some countries, the link between working hours and workload stress may be non-linear, with teachers' quality of life somewhat declining once they work more than 55 hours per week. The time that teachers spend upon certain tasks – most notably marking – is found to be particularly detrimental to their wellbeing, while others (e.g. time spent teaching) have little direct effect. This leads us to conclude that education policymakers and school leaders should aim for a working week that does not exceed around 50 hours during term-time. Hours spent marking, in particular, should be reduced.

The paper now proceeds as follows. A description of the TALIS data is provided in section 2, with an overview of our empirical methodology following in section 3. Results are then presented in section 4, with conclusions then drawn in section 5.

2. Data

The data we use are drawn from the 2018 round of the Teaching and Learning International Study (TALIS). As part of this study, lower-secondary teachers from more than 40 countries completed a questionnaire covering a wide-range of topics about their job. Fieldwork in most countries was completed in the Spring of 2018, though in some Southern Hemisphere countries it was earlier (e.g. in Australia and New Zealand the survey was conducted between September and December 2017).

A two-stage sampling design was used to select teachers and schools, in order to accurately reflect the national teacher population. Schools were first selected with probability proportional to size. Then, within each school, 20 teachers were randomly selected to take part[1]. If a school declined to participate, then a substitute school could take its place. Overall response rates at both the school and teacher level were relatively high, with those for our countries of interest provided in Table 1. Within the international TALIS database (available from https://www.oecd.org/education/talis/talis-2018-data.htm), the survey organisers (the OECD)

[1] If there were less than 20 teachers within a sampled school, then all teachers within the school were asked to participate.

provide a set of sampling and Balanced Repeated Replication (BRR) weights. When applied, these weights fully account for the complex survey design, including making the appropriate adjustment to the estimated standard errors reflecting the clustering of teachers within schools. These weights are applied within our analysis, unless otherwise stated within the notes to the results tables. The final sample size across our five countries of interest is 11,123, with a country-by-country breakdown provided in Table 1. Note that, throughout our analysis, we restrict our attention to full-time teachers only.

<< Table 1 >>

In a smaller number of countries, the TALIS study was also conducted within primary schools. Within our countries of interest, this included England and Australia. The same sampling design was used, with primary school response rates also provided in Table 1. For these two countries, we are hence also able to consider whether the workload-wellbeing relationship differs between primary and secondary staff.

Within one component of the TALIS survey, teachers were asked questions about their hours of work in their last complete working week. The first question simply asked about their total working hours across all tasks:

> During your most recent complete calendar week, approximately how many 60-minute hours did you spend in total on tasks related to your job at this school? [Include time spent on teaching, planning lessons, marking, collaborating with other teachers, participating in staff meetings, participating in professional development and other work tasks. Also include tasks that took place during evenings, weekends and other out of class hours. A 'complete' calendar week is one that was not shortened by breaks, public holidays, sick leave, etc.]

The next two questions were similar, but asked respondents about the time that they spent upon particular tasks:

> Of this total, how many 60-minute hours did you spend on teaching at this school during your most recent complete calendar week?
>
> Approximately how many 60-minute hours did you spend on the following tasks during your most recent complete calendar week, in your job at this school? [Include tasks that

took place during weekends, evenings and other out of class hours. Exclude all time spent teaching, as this was recorded in the previous question.]

Individual planning or preparation of lessons either at school or out of school
Teamwork and dialogue with colleagues within this school
Marking/correcting students' work
Student counselling, supervision and behaviour guidance
Participation in school management
General administrative work
Professional development activities
Communication and co-operation with parents or guardians
Engaging in extracurricular activities
Other work tasks

These variables are the primary covariates of interest throughout our analysis. Descriptive statistics illustrating the distribution of working hours amongst full-time teachers can be found in Table 2. Note that a small number of teachers reported spending an extremely long time upon certain tasks. To limit the impact of such outliers, we cap the top five percent of the distribution for each variable to the 95th percentile. Moreover, any teacher that reported a working week exceeding 84 hours (the equivalent of 12-hour days for seven days a week) has been excluded due to concerns over the reliability of the data they have provided[2].

<< Table 2 >>

Following Allen et al. (2019) and Jerrim and Sims (2019), note that two separate total working hours measures are available within the TALIS data. The first is based upon responses to the single total working hours question as presented above. The second can be derived as the total amount of time that teachers spend upon the 11 separate tasks. One can look at the correlation between these two measures to get an idea of the consistency (and hence reliability) of the information that teachers report about their total hours of work. These are provided, along with the accompanying scatterplots, in Online Appendix A. The two measures tend to be positively correlated (Pearson correlation $\approx 0.5 - 0.7$), though also contain a reasonable amount of noise. Only around half the variance in one of the total working hours measures can be explained by the other. This underlines the importance of considering the impact that such measurement

[2] This restriction means we drop 13% of secondary teachers in Australia, 12% in Alberta-Canada, 10% in England, 9% in New Zealand and 19% in the United States. It also means 13% and 9% of primary teachers are dropped in Australia and England.

error may have upon the results. We shall return to this point in sections 3 (methodology) and 4 (results).

In another part of the TALIS survey, teachers were asked a set of questions about the aspects of their job that causes them stress, and the impact that they believe this has upon their health. Teachers' responses to these questions have then been converted into quasi-continuous scales by the OECD, which form the outcome variables used within our analysis. The first is the "workplace well-being and stress" scale, which is based upon answers to the following questions:

> In your experience as a teacher at this school, to what extent do the following occur:

- I experience stress in my work
- My job leaves me time for my personal life
- My job negatively impacts my mental health
- My job negatively impacts my physical health

With respondents asked to provide one of four response options (not at all, to some extent, quite a bit, a lot). It is therefore a scale that has been designed to capture teachers' subjective views about the impact that their job has upon their general well-being, with a focus upon key issues surrounding mental health (e.g. work-life balance, work-related stress). The scale reliability (Omega coefficient) is above 0.85 within all our countries of interest (see OECD, 2019b: p. 315), though with the second statement ("my job leaves me time for my personal life") having a weaker relationship with the latent factor than the other three.

The second scale we use is more specifically focused upon work-related stress, as captured via the following question:

"Thinking about your job at this school, to what extent are the following sources of stress in your work?"

- Having too much lesson preparation
- Having too many lessons to teach
- Having too much marking
- Having too much administrative work to do (e.g. filling out forms)
- Having extra duties due to absent teachers

The TALIS 2018 technical report (OECD, 2019b, p. 315) again illustrates how the scale reliability is good (between 0.75 and 0.80) within all our countries of interest. The factor loadings were, however, much lower for the statement regarding "extra duties due to absent teachers" suggesting that this question is less central to the construct. Likewise, the factor loading for the statement about administration was also somewhat weaker than for the other three statements, suggesting that this aspect of teacher workload may also be somewhat distinct. For further details, see OECD (2019b). Table 3 provides descriptive statistics for how responses to each well-being question varies across our countries of interest.

<< Table 3 >>

We use these scales as the primary outcomes within our analyses. Yet, in doing so, we must also recognise their limitations. Two key issues emerge. First, the questions capture teachers' self-reported views on the impact of their job upon their well-being and what they *perceive* to be the key sources of stress in their work. This is rather different to some of the measures used in previous studies of teacher well-being (e.g. Bamford & Worth, 2017), such as the General Health Questionnaire, and is clearly subjective. The second (related) issue is cross-national comparability. For instance, the translation and meaning of "stress" could differ substantially across countries and between different cultures. Although the OECD have tested the measurement invariance of the TALIS survey instruments across countries, finding that metric invariance holds for our scales of interest (OECDb, 2019, p. 313), there remain challenges in interpreting responses to such questions across a very wide and diverse pool of nations. This has driven our decision to focus upon five Anglophone education systems (Australia, England, New Zealand, United States and Alberta-Canada) which are linguistically, culturally, politically and economically similar. It also follows a wide body of cross-national comparative research that has specifically focused upon English-speaking countries (e.g. Bradbury et al 2015).

3. Methodology

Total working hours

To begin, we focus upon the relationship between total working hours and teachers' workload, stress and workplace well-being. Following much of the existing literature, our analysis begins by estimating a series of Ordinary Least Squares (OLS) regression models. These are specified as:

$$W_{ij} = \alpha + \beta.H_{ij} + \gamma.D_{ij} + \delta.M_{ij} + u_j + \varepsilon_{ij} \qquad \nabla K \qquad (1)$$

Where:

W_{ij}= One of the teacher well-being scales (as described in the previous section).

H_{ij} = The total number of hours that the teacher reported working during the reference week. Within the main analysis we will use the information on total working hours reported by teachers within a single question. Appendix B then provides alternative results where the measure is based upon the summation of time spent upon different tasks.

D_{ij} = A vector of background characteristics including gender, educational attainment, number of years teaching experience, subject specialism, how teaching qualification obtained.

M_{ij} = A vector of variables capturing the reasons why the individual chose to enter teaching (e.g. reliable income, to contribute to society) and whether teaching was the respondents first-choice career.

u_j = School fixed-effects (i.e. a dummy variable for each school).

ε_{ij} = A random error term.

$_i$ = school i.

$_j$ = school j.

∇K = Indicates that the model is estimated separately within each country, K. Within England and Australia, the model is also estimated separately for primary and secondary teachers.

The parameter of interest from this model is β. This reveals whether teachers who work longer hours have higher or lower levels of well-being than teachers who work shorter hours, amongst those with the same background characteristics and who work within the same school. Final teacher and BRR weights are applied to account for the complex TALIS survey design, including the clustering of teachers within schools. Results will be presented for both the "workplace well-being" and "workload stress" scales described in the previous section. These scales have been standardised to mean zero and standard deviation one within each country, meaning all estimates can be interpreted in terms of effect sizes.

As noted in the previous section, a key limitation of model (1) is that it implicitly assumes that teachers' total working hours are measured without error. In our second set of estimates, we relax this assumption, investigating how our results change under several different assumptions for how reliably total working hours are reported. Throughout this process, measurement error is assumed to be "classical"; in other words, the error is not systematic and is purely random noise. This is implemented via estimation of a structural equation model, where the reliability of total working hours (H_{ij}) is allowed to vary between 0.5 and 1.0. This will, in turn, provide plausible upper and lower bounds on the strength of the association between teachers' workloads and their well-being[3].

Another limitation of the model presented in (1) is that it assumes linearity within the parameters – i.e. that an additional hour of work has the same marginal association with teacher well-being regardless of total workload. Our third set of estimates relaxes this assumption by estimating a model analogous to that presented in equation (1), but now using non-parametric regression, rather than OLS[4]. A key benefit of non-parametric regression is its flexibility, in that it makes no assumption about the appropriate functional form. It is therefore well-suited to our purpose of investigating whether the relationship between teacher workload and well-being is indeed linear, or if there is a certain maximum tolerable length of working week which, if exceeded, has a major negative impact upon teachers' lives.

Time spent upon different tasks

A similar empirical strategy is used to investigate the relationship between the time teachers spend upon certain tasks and their well-being. To begin, we combine some activities together to form the following six workload tasks:

- Teaching (HT)
- Lesson planning and preparation (HP)
- Marking (HM)
- Management and administration (HA)
- Teamworking and professional development (HC)
- Other (HO)

[3] The school fixed effects are not included in our models where we investigate the impact of measurement error upon the results. This is due to models with school fixed effects failing to converge to a unique solution.
[4] The school fixed effects are not included in the non-parametric regression models. Final teacher weights are also not applied.

A matrix illustrating the pairwise correlations between these variables is provided in Online Appendix C (most of the correlations fall below 0.3 in absolute size). The following OLS regression model is then estimated, including each of these variables:

$$W_{ij} = \alpha + \beta_1.HT_{ij} + \beta_2.HP_{ij} + \beta_3.HM_{ij} + \beta_4.HA_{ij} + \beta_5.HC_{ij} + \beta_6.HO_{ij} + \gamma.D_{ij} + \delta.M_{ij} + u_j + \varepsilon_{ij} \quad \nabla K \qquad (2)$$

All variables within this model are defined as before, where HT, HP, HM, HA, HC and HO are variables capturing time spent upon the six different workload tasks (as defined in the bullet points above). The parameters of interest from this model are β_1 to β_6. These illustrate how an extra hour spent upon a given activity (e.g. marking) is related to well-being, conditional upon the time teachers spend upon other activities (e.g. the time they spend teaching), their background characteristics and the school that they teach in. Results from this model will illustrate whether certain aspects of teachers' workloads have a stronger association with the stress caused by their job than others.

Next, we loosen the assumption that the workload variables are measured without error. Following the approach described above, we re-estimate our models six times, with each iteration allowing one of the six tasks to be measured with less than perfect reliability. We thus illustrate how the strength of the association between hours spent upon each workload task and teachers' well-being changes under different assumptions about the magnitude of the measurement error.

To conclude, we re-estimate model (2) within each country using non-parametric regression. This allows us to relax the assumption of linearity and thus establish whether there are certain points where the relationship between time spent upon a certain task and well-being is particularly strong. For instance, is there a maximum amount of marking that teachers can do without their well-being being affected (but, if this level is exceeded, then the impact is substantial)? Results from these non-parametric regression models will shed new light into this important issue.

4. Results

Total working hours

Table 4 begins by presenting OLS estimates of the relationship between teachers' total working hours and their wellbeing. Results are provided for the association with both the workplace

wellbeing and the workload stress scale. In all English-speaking jurisdictions, a modest positive association is observed. For a ten-hour increase in total working hours (roughly the difference between teachers' total working hours in England and the OECD average), workplace stress increases by around 0.15 standard deviations. This pattern is consistent across countries, with few statistically significant cross-national differences in the estimated workload-wellbeing relationship.

<< Table 4 >>

The results in Table 4 are based upon the assumption that teachers' working hours are measured without error. This assumption is relaxed in Figure 1, where we illustrate how the strength of the association between working hours and workload stress changes under different assumptions about how reliably teachers' total working hours are reported. Specifically, the reliability is allowed to vary between 0.5 and 1.0 along the horizontal axis, with the estimated effect of a ten hours increase in total working hours upon workload stress plotted on the vertical axis.

<< Figure 1 >>

Two key points from this graph stand out. First, under plausible assumptions about the reliability of the working hours data, the association between teachers' working hours and their workload stress may be a lot stronger than Table 4 suggests. Take the results for primary teachers in England, for example. The impact of working 10 additional hours per week for this group's workload stress is estimated to be around 0.2 standard deviations, under an assumption of perfect reliability (reliability = 1.0). Yet, if the reliability of the working hours data is assumed to be 0.5 (which is plausible, given the correlations reported in Appendix A) then the strength of this relationship almost doubles, to stand at around 0.35. In other words, since most previous research has failed to consider the potential impact of measurement error, the importance of total working hours for teacher wellbeing may have been underestimated.

The second key feature of Figure 1 is that there is potentially more cross-country variation in the results when the assumption of zero measurement error in the working hours data is relaxed. This is demonstrated by the datapoints for each country sitting tightly together on the right-hand side of the graph (when reliability is 1.0) but are much more spread out on the left (reliability = 0.5). Measurement error in the total working hours data may hence be masking some important international variation in the workload-wellbeing relationship.

Figure 2 turns to a different issue: is the link between teachers' hours of work and their workplace stress / wellbeing really linear? Evidence is presented on this matter based upon results from our non-parametric regressions.

<< Figure 2 >>

Within some populations, a linear relationship between working hours and teacher stress / wellbeing does indeed seem to hold. The clearest examples are for secondary teachers in England (solid blue line) and New Zealand (dashed purple line), where the gradient of the fitted non-parametric regression line is broadly constant across the range of working hours plotted (40 to 65 hours per week). In other words, for these groups, each additional hour of work is related to the same decline in workplace wellbeing (and increase in workload stress), regardless of existing total hours of work.

Yet the same is not true elsewhere. Take, for instance, primary teachers in England and secondary teachers in Australia. The non-parametric regression line is quite flat between 40 and 50-55 hours per week (which roughly corresponds to the bottom half of the weekly working hours distribution for full-time teachers with this countries). Hence increasing working hours for those teachers working within this range has little association with their levels of stress and wellbeing in the workplace. However, once working hours exceed this level, then each additional hour of work has a strong negative association with teachers' quality of life. This is demonstrated by the sharp turn in the non-parametric regression lines for England (primary) and Australia (secondary) between around 50 ad 65 hours per week.

Together, the evidence presented in Figure 2 leads us to reach three important conclusions. First, one should not assume that the teacher workload-wellbeing relationship is necessarily linear. Despite this assumption being implicitly made in much of the existing literature, we find some evidence of important non-linearities occurring within some countries. Second, the nature of the association between working hours and wellbeing may differ across countries and, indeed, between different groups within a country (e.g. in Figure 2 there is a clear difference between primary and secondary teachers in England). Hence results from previous studies, focusing upon a single group within one specific national setting, may not generalise to other education systems (i.e. results may have relatively weak levels of external validity).

Finally, in terms of policy and practise, a strong case can be made for reducing inequality in full-time teachers' working hours in some jurisdictions. For instance, Figure 2 suggests that

full-time primary teachers in England who work 40 hours per week could increase this to 45 or even 50 hours with little effect upon their workload stress and wellbeing. At the same time, a reduction of five or ten hours amongst those teachers who currently work 60 or more hours per week might potentially lead to an appreciable increase in this group's quality of life. School leaders (and education policymakers) who are able to share the workload burden equally amongst staff may hence have a less stressed and healthier workforce than those who lead schools where the distribution of workload is very unequal.

Time spent upon different tasks

Our analysis now turns to the amount of time that teachers spend upon different tasks. The OLS regression results are presented in Table 5, with all estimates conditional upon the time spent on the other remaining activities (e.g. the association between time spent teaching and workload stress is conditional upon the time spent upon marking, lesson planning etc). The shading of cells should be read horizontally, with red (green) indicating a stronger (weaker) negative association with workload stress. Analogous results using the workplace wellbeing scale (rather than the workload stress scale) can be found in Appendix D.

<< Table 5 >>

Across all English-speaking jurisdictions, there are two clear aspects of the job that are strongly associated with teachers' workload stress. The first is marking, with each additional hour spent on this task associated with around a 0.06 standard deviation change in the stress that teachers feel at work. The direction and magnitude of this effect is similar across all the education systems considered; marking, however it is done, always seems to be a stressful part of the job. Lesson planning and preparation is the other aspect of teachers' jobs that consistently has a negative association with workload stress. Across the countries that we consider, each additional hour spent upon lesson planning raises workload stress by 0.034 standard deviations, ranging from a high of 0.053 standard deviations in Alberta to a low of 0.018 for primary teachers in Australia.

However, outside of these two areas, the association between time spent upon the other tasks and their workload stress is rather weak. For instance, after accounting for any additional time that must be allocated to lesson preparation and marking, each additional hour spent teaching has little correlation with our outcome measures. Take primary teachers in England. A substantial 10 hour increase in their teaching load is associated with just a 0.07 standard

deviation increase in their level of workload stress (conditional upon this not increasing the time that they spend on other tasks). This suggests that it is not teaching hours per se that causes teachers stress, but rather the auxiliary tasks (marking, planning) that come with it. Interestingly, the effect of each additional hour allocated to management and administration is essentially zero across most of the countries considered. Meanwhile, additional time allocated to collaboration with colleagues and CPD may actually lead to a reduction in workload stress (effect size = 0.03 for each additional hour), though in many countries this association does not quite reach statistical significance at conventional levels. As indicated by the similar pattern of shading across cells, very similar patterns emerge across countries and between primary and secondary staff.

Figure 3 illustrates how the link between the different tasks and workload stress varies under different assumptions about the reliability of the working hours data. Results are presented for secondary teachers in England as an illustration, with those for other countries presented in Appendix E. For management/administration, teamwork/CPD and other, the impact that measurement error has upon the results is trivial. Even when the reliability of the working time data is assumed to be quite low, the time spent upon these tasks is only weakly associated with workload stress. For instance, the effect of each additional hour of management /administration and other tasks is only around 0.01 standard deviations when the reliability of the working hours data is set at 0.6. This helps strengthen the case that the amount of time that teachers spend upon such tasks is largely unrelated to their wellbeing.

<< Figure 3 >>

On the other hand, it seems that the link between hours spent teaching, marking and lesson preparation and their workload stress may be somewhat underestimated due to measurement error. This is illustrated by the fact that the estimated effect sizes for these three tasks increase somewhat when the reliability of measurement is assumed to be lower. The clearest example is for the time spent upon teaching, where the estimated effect doubles when moving from an assumption of 'perfect' reliability (1.0) to an assumption where the working hours data are measured with quite a substantial degree of error (reliability = 0.5). There is hence a case that the association between marking, lesson planning, teaching and workload stress might be somewhat stronger than the OLS results presented in Table 5 suggest.

To conclude, Figure 4 turns to our investigations of non-linearities. Estimates are presented for the three tasks where a relationship with workload stress was found using OLS: (a) teaching;

(b) lesson planning and (c) marking. Although a broadly linear relationship can be observed for most tasks in most countries, there are some important exceptions. This includes time spent upon lesson planning in the United States and time spent upon marking for secondary teachers in Australia, for example. However, the most striking case is secondary teachers in England. Our discussion therefore focuses upon the results for this group.

<< Figure 4 >>

With respect to time spent upon teaching, workload stress for secondary teachers in England reaches its peak at 23 hours per week, after which point workload stress declines. One potential explanation for this result is that those teachers who are required to teach for a greater number of hours are relieved from some other duties. On the other hand, the link between lesson planning and workload stress for secondary teachers in England is broadly linear up to seven hours per week, but then tails off. In other words, the first few hours spent upon lesson planning have the greatest marginal impact upon the stress suffered by this group. Finally, there is a big jump in workload stress for secondary teachers in England who spend between one and five hours marking students work each week. The estimated non-parametric regression line is then flat between five and ten hours, suggesting that each additional hour of marking within this range has little association with workload stress. However, once this ten-hour threshold for marking is exceeded, then workload stress increases exponentially. This suggests that 10 hours of marking per week represents a red line for secondary teachers in England that should not be crossed. As almost a quarter of full-time secondary teachers in England currently spend ten hours per week or more on marking[5], this represents a priority for workload reduction in England.

5. Conclusion

In many Western education systems, teachers are expected to work long hours each week (Jerrim & Sims, 2019), particularly during term-time. There is increasing concern amongst policymakers, and the wider education community, about the impact that this is having upon teachers' work-life balance and their general well-being (Copper-Gibson Research 2018). Indeed, the long and stressful working hours associated with teaching are thought by many to be causing a recruitment and retention crisis in the profession (Foster, 2019). A prime example is England, a country where teachers work particularly long hours (Allen et al., 2019), spend a

[5] Authors' calculations based upon the TALIS 2018 sample used in this paper.

comparatively long time upon burdensome non-teaching tasks such as marking, have low-levels of teacher well-being and where many school-leaders believe that a lack of suitably qualified staff is hindering instruction within their school (Jerrim & Sims, 2019). Finding out more about how teachers' working hours are affecting their well-being has hence become an issue of major education policy interest, with further insights needed so that many of the challenges facing the teaching profession can be resolved.

Yet, despite the widespread interest in this issue, some clear gaps within the existing academic evidence base remain. For instance, is it really the total amount of time that teachers spend working that causes them stress? Or is it the time that they spend upon particularly unenjoyable tasks (e.g. marking)? Might the relationship between workload and well-being be non-linear, with the marginal impact of each additional hour depending upon how much work teachers already do? Could existing evidence on the strength of the workload-wellbeing relationship be underestimated, given that teachers' working hours are likely to be measured with error? And what about the external validity of results – does the amount of time spent upon certain tasks have a negative impact upon teacher stress and wellbeing within some education systems, but not in others?

Using recently released data from TALIS 2018, this paper has attempted to provide new evidence on such issues, demonstrating how the association between workload and wellbeing varies across five predominately English-speaking jurisdictions. Our results illustrate how the link between teachers' working hours and their workplace wellbeing may not be linear (at least in some countries) and, after taking into account the potential impact of measurement error, may be stronger than previously thought. The two aspects of teachers' jobs that leads to the greatest increase in workload stress are lesson planning and marking; each additional hour spent upon the latter is associated with a 0.06 standard deviation increase in stress in the workplace. This is in contrast to other aspects of the job, such as time spent teaching and working with colleagues / professional development, which seem to have little direct effect upon teachers' quality of working life. Together, the evidence points towards a need to aim for a term-time working week for teachers of no more than around 50 hours - similar to the maximum allowed under the European Working Time Directive. This reduction in total working hours should be mainly achieved via reductions in marking and lesson planning.

It is of course important that these findings are interpreted considering the limitations of this study. Three issues stand out. First, the outcome measures are subjective, and in many ways

capture teachers own views of what causes them stress and affects their well-being in the workplace. Future data collections and subsequent research should investigate whether similar findings emerge using other widely used measures of workplace stress and mental health, such as the General Health Questionnaire (Goldberg & Hillier, 1979). Second, a key strength of this paper is that we have explicitly considered the potential impact that measurement error may have upon our results. Our investigations have, however, been conducted under an assumption that such measurement error is "classical" (i.e. random noise). It is well-known that dealing with more complex forms of measurement error is a much greater challenge (Black, Berger, & Scott, 2000), particularly when little is currently known about whether teachers with certain characteristics are more likely to over or under report their hours of work. This paper should hence be considered a first step towards addressing the problem of measurement error within this literature, with future data collections needed (possibly gathering time-use diary data from teachers) to provide further information on this issue. Finally, as per most empirical work in this area, our estimates refer to conditional associations and may not be capturing cause and effect. Future studies, possibly drawing upon longitudinal data, may be of value in strengthening the evidence of whether teachers' hours of work are causally related to their workload stress and workplace wellbeing.

Despite these limitations, we believe that our findings have important implications for education policy. There are two clear areas where reducing teachers' workloads would help reduce stress: lesson preparation and marking. With respect to the former, perhaps the easiest thing that policymakers can do is dramatically reduce examination, curriculum and inspection reforms – all of which create new work for teachers, who are forced to change lessons, materials and teaching styles as a result. Indeed, in 2019, the Department for Education in England committed to "a period of greater stability in curriculum, qualifications and assessment" (DfE, 2019, p. 15). Given our findings on the relationship between stress and both lesson planning and overall workload, many teachers in other countries would likely a similar pledge.

The time allocated to marking may be harder to reduce, though it is noteworthy how the amount of time spent marking in several English-speaking countries (e.g. both primary and secondary teachers in England) is above international averages (OECD, 2019a). A recent review found a lack of good research on the impact of written marking on pupil learning (Elliott et al., 2016). However, it found that there was evidence to suggest that acknowledgement marking, awarding grades for every piece of work and marking without providing time for pupils to consider the feedback are all unlikely to help pupils – thus, "school should mark less… but better." (p. 5).

Consequently, it is likely that many schools could reduce teacher stress by reducing the amount of marking they require teachers to do, without it having a detrimental effect on pupil learning.

References

Allen, R.; Benhenda, A.; Jerrim, J. and Sims, S. 2019. New evidence on teachers' working hours in England. An empirical analysis of four datasets.

Bamford, S. and Worth, J. 2017. 'Is the grass greener beyond teaching?' Accessed 01/06/2018 from https://www.nfer.ac.uk/teacher-retention-and-turnover-research-research-update-3-is-the-grass-greener-beyond-teaching

Black, D.; Berger, M. and Scott, F. 2000. Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association* 95(451): 739-748.

Bradbury, B.; Corak, M.; Waldfogel, J. and Washbrook, E. 2015. *Too Many Children Left Behind: The U.S. Achievement Gap in Comparative Perspective*. Russell Sage Foundation, New York.

Bradbury, A., & Roberts-Holmes, G. (2018). *The datafication of primary and early years education: Playing with numbers*. Routledge.

Cooper-Gibson Research. 2018. Factors affecting teacher retention: qualitative investigation. Department for Education Research Report. Accessed 27/08/2019 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/686947/Factors_affecting_teacher_retention_-_qualitative_investigation.pdf

Department of Education [DfE] (2019). Teacher recruitment and retention strategy. Accessed 27/01/20 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786856/DFE_Teacher_Retention_Strategy_Report.pdf

Elliott, V., et al. (2016). *A marked improvement? A review of the evidence on written marking*. Education Endowment Foundation. Accessed 27/01/20 from https://educationendowmentfoundation.org.uk/public/files/Presentations/Publications/EEF_Marking_Review_April_2016.pdf

Fernet, C., Austin, S., Trépanier, S. G., & Dussault, M. (2013). How do job characteristics contribute to burnout? Exploring the distinct mediating roles of perceived autonomy, competence, and relatedness. *European Journal of Work and Organizational Psychology*, *22*(2), 123-137.

Fernet, C., Guay, F., Senécal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education*, *28*(4), 514-525.

Foster, D. 2019. Teacher recruitment and retention in England. House of Commons briefing paper 7222. Accessed 17/07/2019 from https://researchbriefings.files.parliament.uk/documents/CBP-7222/CBP-7222.pdf

Goldberg, D and Hillier, V. 1979. A scaled version of the General Health Questionnaire. *Psychological Medicine*. 9(1):139–145.

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of school psychology*, *43*(6), 495-513.

Jerrim, J., & Sims, S. 2019. *The Teaching and Learning International Survey (TALIS) 2018*. London: Department for Education.

Kokkinos, C. M. (2007). Job stressors, personality and burnout in primary school teachers. *British Journal of Educational Psychology*, *77*(1), 229-243.

Kinman, G., Wray, S., & Strange, C. (2011). Emotional labour, burnout and job satisfaction in UK teachers: The role of workplace social support. *Educational Psychology*, *31*(7), 843-856.

Lam, B. H., & Yan, H. F. (2011). Beginning teachers' job satisfaction: The impact of school-based factors. *Teacher Development*, *15*(3), 333-348.

National Education Union. 2018. Teachers' workloads. Accessed 27/08/2019 from https://neu.org.uk/policy/teachers-workload

OECD. 2019a. TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners. Paris: OECD. Accessed 27/08/2019 from http://www.oecd.org/education/talis/

OECD. 2019b. TALIS 2018 Technical Report: Preliminary version. Paris: OECD. Accessed 27/08/2019 from https://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf

Perryman, J. and Calvert, G. 2019. What motivates people to teach, and why do they leave? Accountability, performativity and teacher retention *British Journal of Educational Studies* DOI: 10.1080/00071005.2019.1589417

Selwyn, N., Nemorin, S., & Johnson, N. (2017). High-tech, hard work: An investigation of teachers' work in the digital age. *Learning, Media and Technology*, *42*(4), 390-405.

Skaalvik, E. M., & Skaalvik, S. (2009). Does school context matter? Relations with teacher burnout and job satisfaction. *Teaching and teacher education*, *25*(3), 518-524.

Table 1. Sample sizes and response rates across selected countries

	School response rate		Teachers in participating schools	Full-time teacher sample size
	Before replacement	After replacement		
Australia (primary)	49%	74%	76%	1,743
Australia (secondary)	49%	75%	78%	2,376
Alberta-Canada	49%	60%	84%	748
England (primary)	74%	86%	85%	1,318
England (secondary)	73%	82%	84%	1,631
New Zealand	63%	79%	80%	1,589
United States	60%	77%	90%	1,718

Notes: Weighted response rates reported. See OECD (2019b: Appendix G) for further details.

Table 2. Average working hours spent upon different workload tasks across countries

Country	Total	Teaching	Planning	Marking	Management + admin	Teamwork + CPD	Other
England - primary	54.3	24.0	7.1	5.8	5.4	4.5	5.8
England - secondary	51.8	20.1	7.2	6.1	5.5	3.9	7.2
Alberta	51.1	27.7	6.7	4.5	2.8	3.5	5.8
Australia - primary	50.1	24.4	7.4	3.1	4.5	4.7	5.9
USA	50.1	27.3	6.3	4.5	2.4	3.6	9.0
New Zealand	49.3	19.7	6.4	4.4	5.7	4.9	7.2
Australia - secondary	49.2	19.5	6.9	4.6	5.7	5.0	6.9

Notes: Sum of each task does not equal the total, as information is drawn from different questions.

Table 3. The distribution of teacher well-being measures across selected countries

	Australia - primary	Australia - secondary	Alberta	England - primary	England - secondary	New Zealand	USA
Workplace wellbeing							
Experience stress	58%	59%	60%	66%	71%	61%	56%
Time for personal life	28%	33%	27%	23%	22%	30%	48%
Impacts mental health	26%	24%	25%	28%	35%	27%	16%
Impacts physical health	20%	21%	18%	22%	27%	21%	14%
Workload stress							
Too much prep	40%	32%	32%	45%	40%	30%	30%
Too many lessons	30%	25%	26%	29%	40%	24%	22%
Too much marking	33%	45%	43%	54%	71%	51%	38%
Too much admin	50%	57%	31%	59%	66%	67%	32%
Extra duties	14%	24%	12%	21%	31%	23%	11%

Notes: Figures refer to percentage of teachers who selected either “quite a bit” or “a lot”.

Table 4. OLS estimates of the relationship between total working hours and teacher well-being

	Workplace well-being		Workload stress	
	Effect size	**SE**	**Effect size**	**SE**
England – secondary	0.20*	0.03	0.13*	0.03
England – primary	0.19*	0.03	0.18*	0.03
Alberta	0.18*	0.05	0.16*	0.05
New Zealand	0.16*	0.03	0.14*	0.03
Australia – primary	0.15*	0.03	0.13*	0.03
Australia – secondary	0.14*	0.02	0.13*	0.03
USA	0.10*	0.04	0.18*	0.04

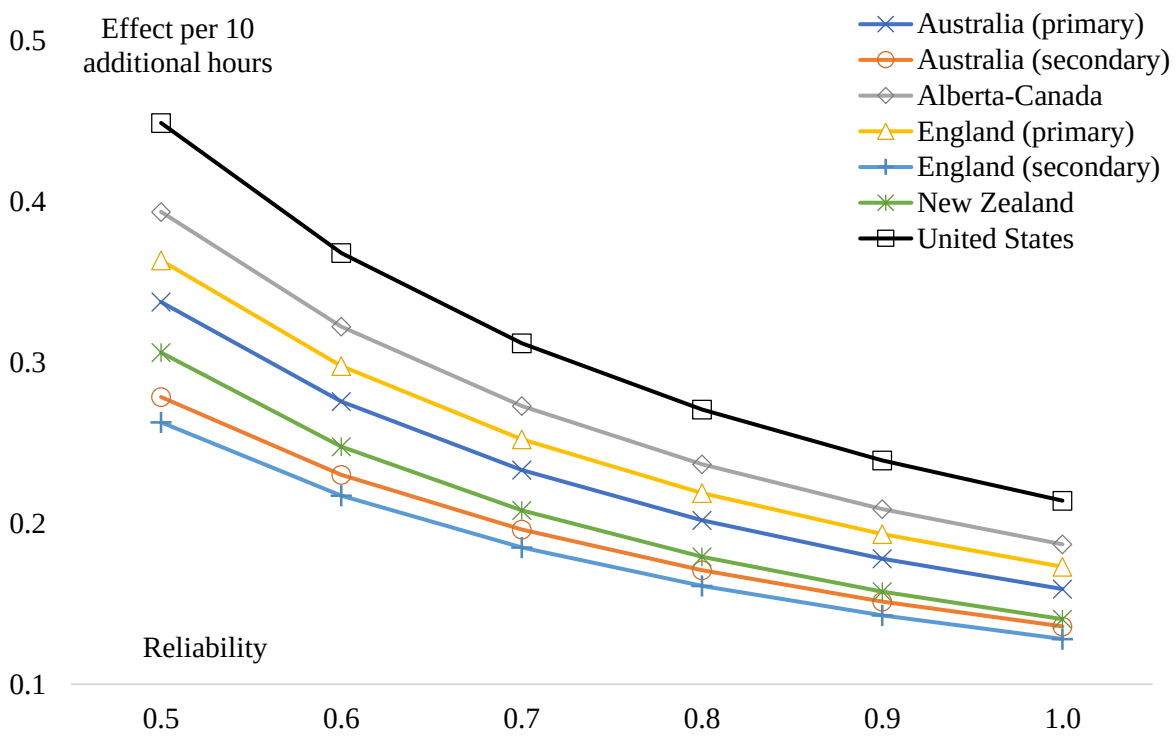
Notes: Figures refer to effect size change in the well-being / stress scale per ten-hour increase in total working hours. * indicates statistical significance at the five percent level.

Table 5. OLS estimates of the relationship between time spent upon marking, management + administration and workload stress

Country	Teaching	Preparation	Marking	Management & admin	Teamwork & CPD	Other
Alberta	0.012*	0.053*	0.066*	0.023	-0.050*	0.004
Australia - primary	0.015*	0.018*	0.064*	-0.008	-0.021	0.016*
Australia - secondary	0.012*	0.042*	0.051*	0.003	-0.016	0.005
England - primary	0.007	0.038*	0.060*	-0.008	-0.027*	0.019*
England - secondary	0.029*	0.030*	0.046*	-0.005	-0.013	0.006
New Zealand	0.019*	0.027*	0.062*	0.008	-0.032*	0.000
USA	0.004	0.034*	0.068*	0.003	-0.045	-0.007
Average	**0.014**	**0.034**	**0.059**	**0.002**	**-0.029**	**0.006**

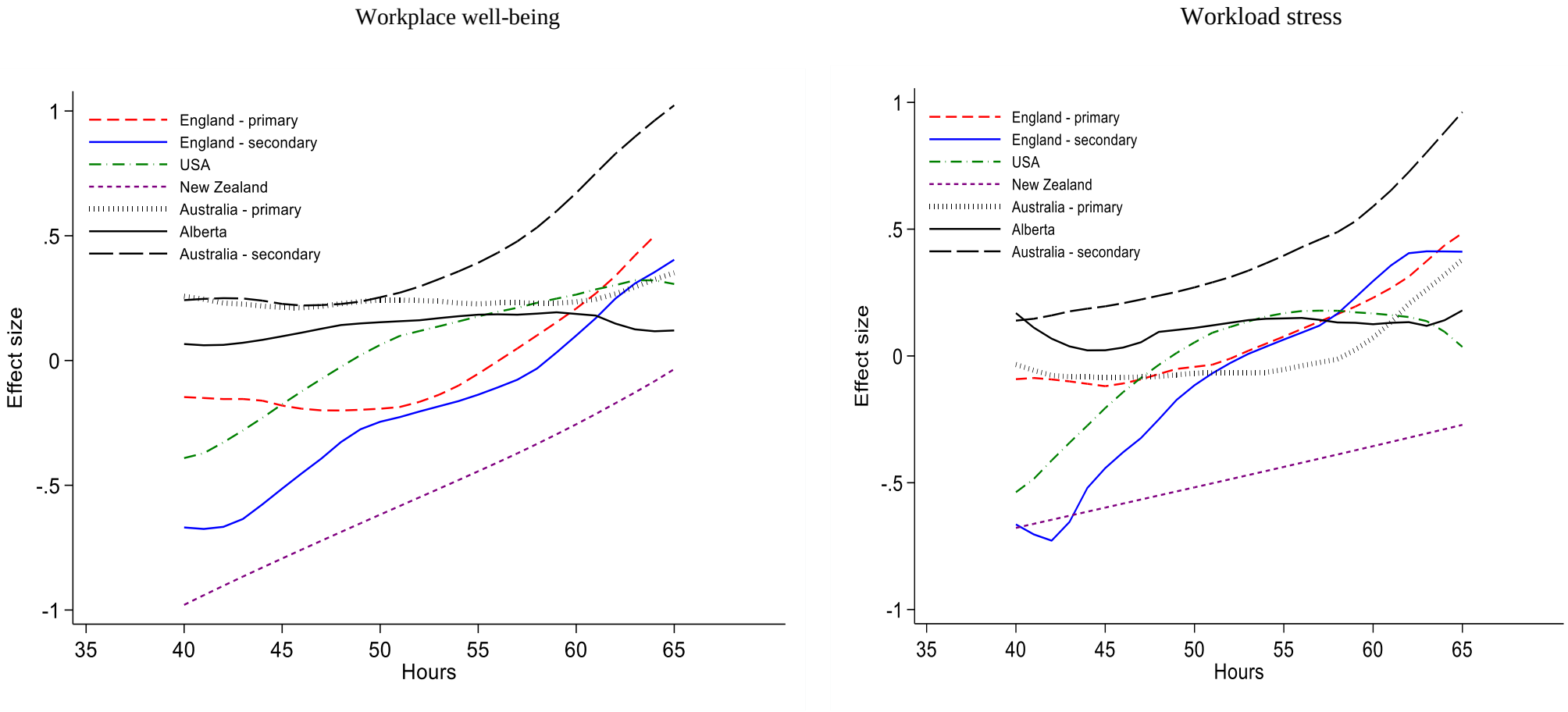
Notes: Figures refer to effect size change in the well-being / stress scale per one hour increase in each task. Shading should be read horizontally. Redder shading indicates a greater negative association with workload stress, green shading indicates a zero or negative association with workload stress. A * indicates significantly different from zero at the five percent level.

Figure 1. The relationship between total working hours and workload stress under different assumptions about the reliability of working hours measurement



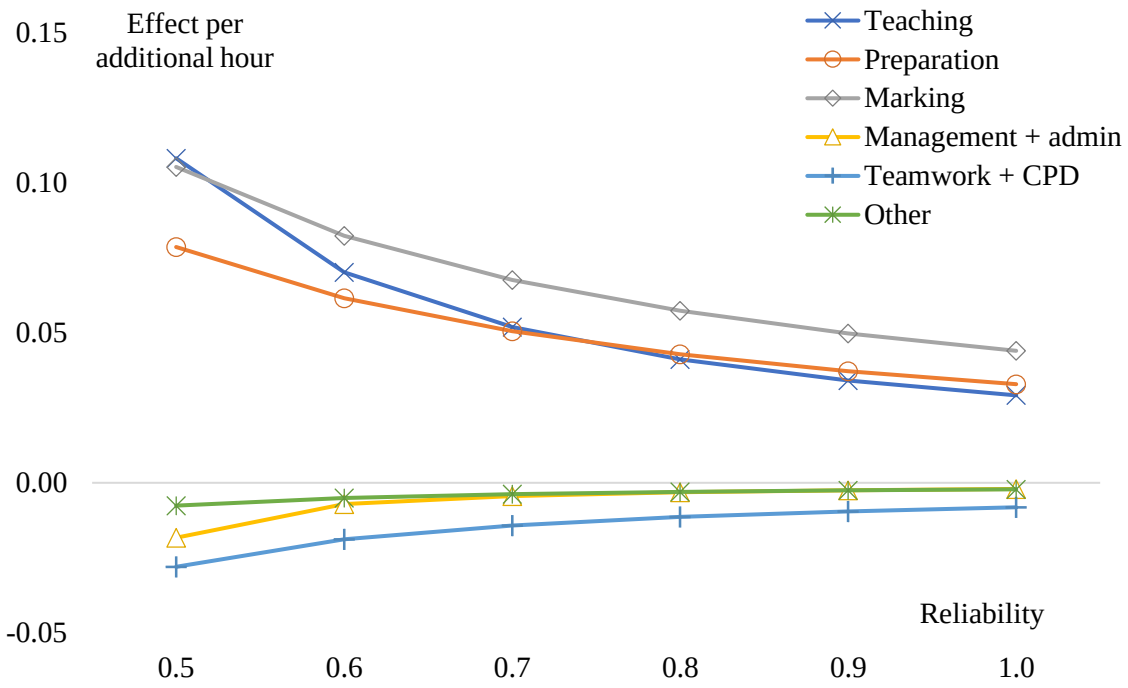
Notes: See Appendix D for analogous results for the workplace well-being scale. School fixed effects have been excluded from the model. Weights not applied.

Figure 2. Non-parametric regression estimates of the association between total working hours and workplace well-being and stress



Notes: Weights not applied. Models exclude the school fixed effects. Results shown for teachers working between 40 and 65 hours per week (approximately the 10th and 90th percentile of the working hours distribution). Results based upon a female teacher, with number of years teaching experience set to the national average, who holds an ISCED Level 6 (postgraduate) qualification, whose motivations for entering teaching were similar to the national average, but for whom teaching was not their first-choice career.

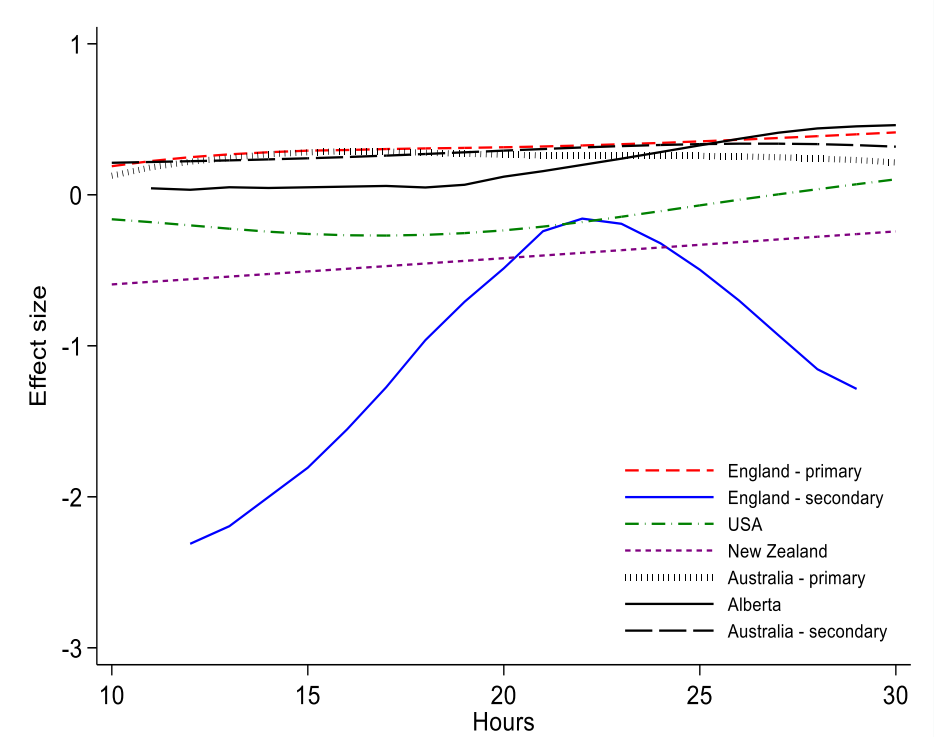
Figure 3. The association between the time spent upon different tasks and workload stress. Estimates for England (secondary) under different assumptions about reliability.



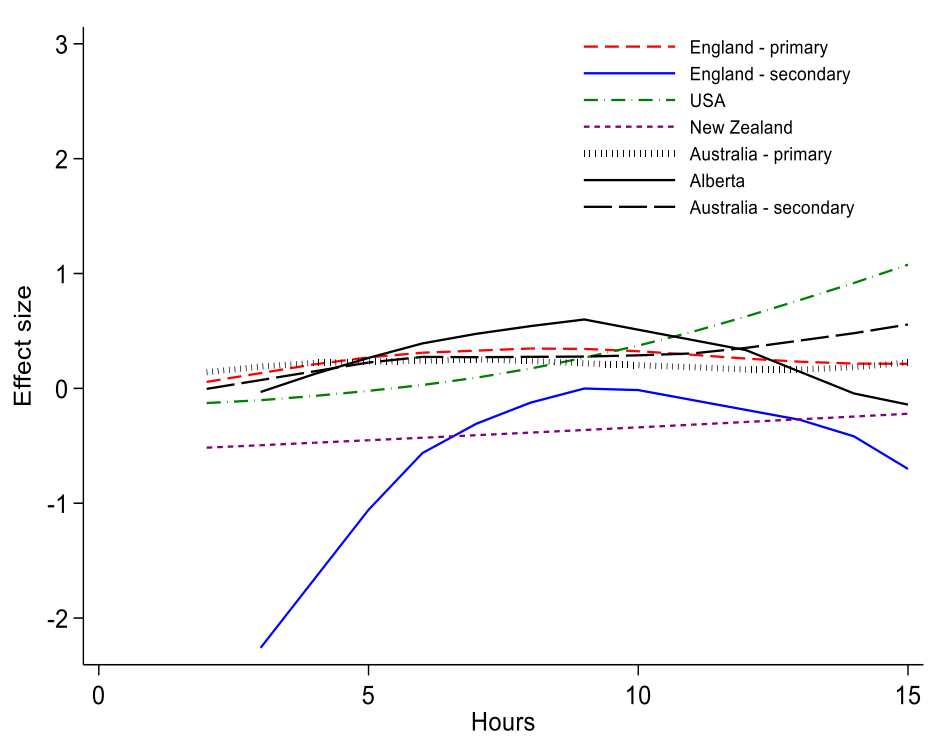
Notes: Estimates based upon six structural equation models (one for each workload measure). Within these models, the reliability of the workload measure of interest (e.g. marking) is allowed to vary between 0.5 and 1.0. The other five workload measures included in the model (e.g. teaching, preparation, management and administration, teamwork and CPD, other) are assumed to be measured without error.

Figure 4. Non-parametric regression estimates of the association between the time spent upon different tasks and workload stress.

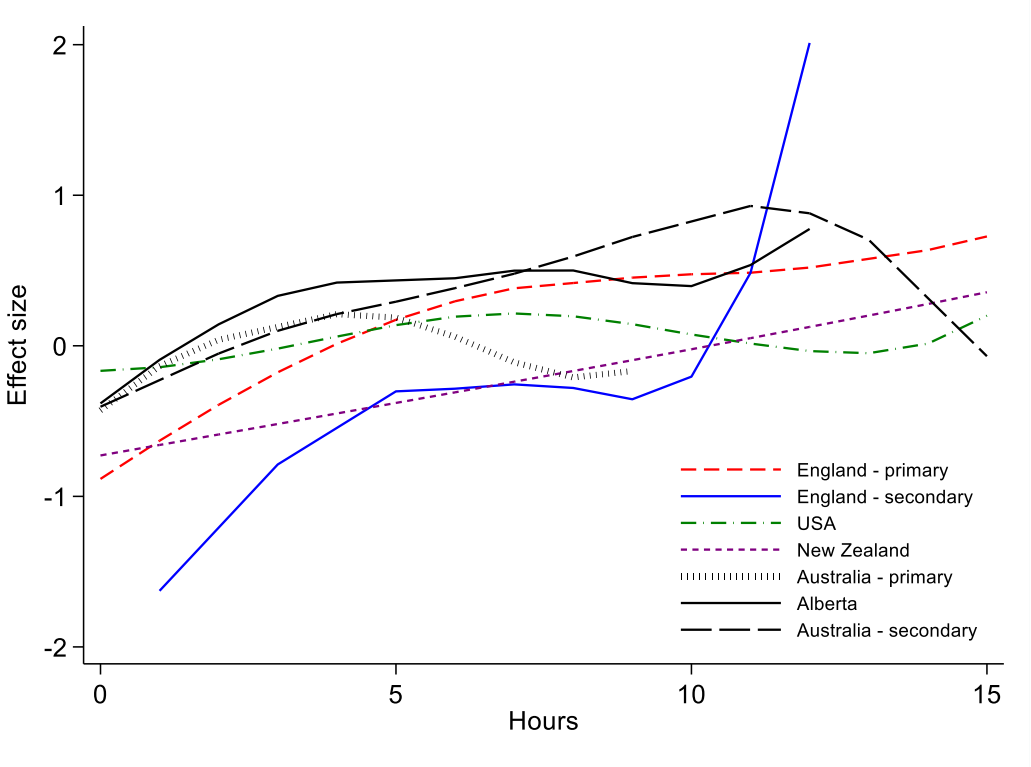
(a) Teaching



(b) Lesson planning



(c)Marking



Notes: See Appendix D for analogous results for the workplace well-being scale. Weights not applied. Model estimated excluding the school fixed effects. Results based upon a female teacher, with number of years teaching experience set to the national average, who holds an ISCED Level 6 (postgraduate) qualification, whose motivations for entering teaching were similar to the national average, but for whom teaching was not their first-choice career.