

High stakes assessments in primary schools and teachers' anxiety about work

John Jerrim
Rebecca Allen
Sam Sims

October 2022

High-stakes assessments – where the results have important consequences for teachers, pupils and/or schools – are a common feature of many education systems across the world. One argument often made against their use, however, is that they have a negative impact on wellbeing across the education sector. This includes those teachers who hold most responsibility for the results. We present new evidence on this matter by examining how the Year 6 Statutory Assessment Tests (SATs) conducted in England's primary schools are linked to how anxious teachers feel about work. Drawing on unique panel data from around 1,000 primary school teachers, we illustrate how the SATs are associated with a short, sharp shock to teachers' anxiety levels during the week the tests take place. Yet there is little evidence that those most exposed to the pressures of SATs suffer from especially prolonged periods of high anxiety levels. We thus conclude that, although there may be other reasons to lower the stakes attached to primary school assessments, the benefits for teachers' anxiety levels are likely to be marginal.

Key Words: Assessment, SATs, Key Stage 2, anxiety, teachers, accountability.

Acknowledgements: The Nuffield Foundation is an independent charitable trust with a mission to advance social wellbeing. 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 expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org. We are grateful for their support. Helpful comments have been received on the draft from our project advisory group, whom we would like to thank for their input and support.

Contact details: John Jerrim, UCL Social Research Institute. J.Jerrim@ucl.ac.uk

1. Introduction

High-stakes assessments are a prominent feature of many education systems. These measure young people's academic competencies in key areas, with the results potentially having important consequences for pupils, schools, and/or staff. Data from such assessments are increasingly used to hold educators to account. This includes via the publication of school "league tables", performance-related pay initiatives and threats to change the school management structure in response to results. Although high-stakes tests are now common in secondary schools, they are increasingly being used in the primary sector as well. England – the empirical setting for this paper – is a prime example. Here, 10/11-year-olds sit Key Stage 2 Statutory Assessment Tests (SATs) at the end of primary education, with the results potentially having important consequences for schools, leaders, teachers and – to a lesser extent – individual pupils.

The proponents of high-stakes tests argue that assessments such as England's SATs help to monitor and maintain standards in schools, leading to improvements in achievement (Saminsky 2011). Yet it is widely recognised they can have unintended negative impacts as well. This includes increasing teacher workload (Perryman and Calvert 2020), narrowing the curriculum to focus on only those subjects tested (Berliner 2011), incentivising schools to "cream-skim" the best pupils (West and Pannell 2006) and excluding those deemed at greatest risk of failure (Ofsted 2019). Schools may then not always make decisions in the best interest of pupils (Ofsted 2019). Such issues have led some to call for high-stakes assessments to be scrapped (National Education Union 2019).

One particularly pertinent topic that has recently come to the fore is how high-stakes assessments impact mental health and wellbeing in the education sector. Although much work in this area has focused on pupils, concern has been shown about the effects of high stakes testing on staff as well. For instance, by encouraging teaching-to-the-test, and subsequent narrowing of the curricula, it has been argued that teachers may become less satisfied in their jobs (Smith and Holloway 2020). Given they and/or their school may suffer negative consequences if the results are poor, high-stakes tests may lead teachers to become more anxious or stressed about work. It may also increase teachers' workloads – and thus subsequently reduce their wellbeing – as well. Such arguments have been used as a further

platform to support the case for ending high-stakes assessments, particularly in primary schools (Bradbury 2019).

Yet much of the previous work on the link between high stakes assessments and staff wellbeing has focused on the secondary education sector. For instance, using data from a survey of 145 Texan teachers – most of whom taught middle or high school pupils – Gonzalez et al (2017) found there to be a “*significant mean difference in overall stress among high school teachers who taught high-stakes subject matter compared with their non-high-stakes subject matter counterparts*”. Focusing on assessment reforms in Ontario, Hargreaves (2020) argues that even “mid-stakes” tests are perceived to have a range of negative consequences, including leading to emotional ill-health amongst both students and staff. Qualitative interviews with 13 middle school mathematics teachers by Demir and Keleş (2021) found high-stakes testing to be associated with teachers’ anxiety and stress levels, which in-turn impacted their instruction. Using data from a survey of 18 elementary school teachers, Gunn, Al-Bataineh and Abu Al-Rub (2016) detailed how most teachers feel pressured by high-stakes tests, with this stemming from demands within the local community and by managers, as well as individuals putting pressure on themselves. In a survey of more than 900 teachers about the Texas Assessment of Academic Skills, Reese, Gordon and Price (2004) conclude that high stakes testing leads to “*the entire educational system [becoming] stressed*”. In a survey of 300 teachers in Georgia, Brockmeier et al (2014) argue that teachers perceive the pressures of high stakes testing to be closely related to their stress levels at work. Discussing the evidence on high stakes testing and teacher stress, Mathison and Freeman (2006) suggest that such assessment regimes are “*likely to have unhealthy effects and lead to stressed teachers who struggle along, get sick, or leave the profession*”. Turning to England, a survey commissioned by the National Education Union claimed that 90% of teachers believe the end-of-primary SATs negatively impacts the mental health of teaching staff (TES 2018). Similarly, in a survey completed by 188 teachers in England conducted by Bradbury (2019) and funded by the interest group More Than a Score, 92% agreed or strongly agreed that SATs have a negative impact on teachers’ wellbeing.

Despite the insights provided by the aforementioned studies, there continues to be gaps in our knowledge about the relationship between high stakes tests and the mental wellbeing of teachers. Four notable issues stand out. First, much existing research is small-scale and cross-sectional. There is a notable dearth of large-scale longitudinal quantitative evidence into how high-stakes tests are associated with teachers’ wellbeing. Second, there has been little attempt

to investigate how the wellbeing of teachers who are most affected by such tests compare against a reasonable “comparison group” (teachers less likely to be affected by such tests). Third, no existing study investigates whether staff become more anxious about work around the time that high-stakes tests take place, as the “threat” from them draws near (Lotz and Sparfeldt 2017). Finally, in general, less attention has been paid to the link between high stakes tests and staff wellbeing in primary schools than in secondary schools.

This paper provides new insights into such issues. Using longitudinal data gathered from around 1,000 teachers in England, we examine how work-related anxiety fluctuated over the 2021/22 academic year. Particular attention is paid to Year 6 teachers – the group most affected by the high-stakes SATs regime in England – in comparison to those providing instruction to other primary school classes (e.g. teachers of Year 3/5 pupils who do not sit any national assessments). This includes an analysis of how the work-related anxiety of these groups compare in the build-up to - and during - when the SATs take place. In doing so, we provide unique new insights into how high-stakes assessment is linked to the work-related anxiety of staff in primary schools. Specifically, the following research questions are addressed:

- *Research question 1. Do Year 6 teachers – whose pupils sit the high-stakes SATs - have higher levels of work-related anxiety than primary teachers providing instruction to other school year groups (whose pupils do not sit any national assessments)?*
- *Research question 2. Does the work-related anxiety of Year 6 teachers increase in the build-up to - and during - SATs test week? How does this compare to teachers providing instruction to other school year groups?*

The paper now proceeds as follows. Section 2 provides further background into why teachers wellbeing at work may be impacted by the presence of high-stakes tests. A description of our data and empirical methodology follows in section 3. Results are presented in section 4, with conclusions then drawn in section 5.

2. Why might high-stakes assessments impact the wellbeing of teachers?

The use of test scores by authority figures within the school

Senior leaders, managers and governors are often able to access high-stakes test results at the class – and hence individual teacher – level. Although they may not be designed - and indeed may be ill-suited - for this purpose, authority figures within schools may use such data to make judgements about the performance of individual teachers. This may in-turn lead to increased anxiety about such tests if teachers feel they are being monitored (Page 2017) – particularly if negative consequences are attached. For instance, around a quarter of teachers in England report their pay progression to be tied to their class achieving certain examination results (Teacher Tapp 2019a), while for some it will also form part of their annual appraisal (Neumann et al 2016).

League Tables

In some countries – including England – the government publishes “league tables” facilitating comparisons of performance in high-stakes tests across schools. To aid interpretation, all schools in England are divided into one of five bands depending on the results (from well below average to well above average). This may cause stress for those directly involved in such tests (e.g. teachers, school leaders) whose “performance” will be publicly known. Given the small number of classes in many primary schools, this may pose a particular problem. For instance, in schools with a single-form entry, one would know the average scores of pupils taught by the Year 6 teacher. Proponents of high-stakes test accountability argue that such open sharing of information will incentivise schools and teachers to improve pupil achievement (Supovitz 2009) – at least in the areas under examination. Yet it is also likely to cause anxiety amongst staff – particularly those teaching the pupils taking the tests – with any slip in performance (such as SATs scores moving from “average” to “below average”) becoming widely known.

Consequences for their colleagues

Teachers may feel more anxious about high-stakes tests if the results also have consequences for their colleagues. Previous research has found individuals may become anxious if they feel they may let others down (Tucker and Horton 2019). For instance, teachers may believe that poor performance in a high-stakes test could trigger a school (Ofsted) inspection, or that this could lead to a downgrade in the school’s inspection judgement. This may then also put

pressure of them from their colleagues as well. Those holding most responsibility for the high-stakes tests within their school will be aware of such consequences, and which may thus impact their wellbeing as a result.

Living up to the expectations of others

Given the importance of high-stakes tests, it has been reported that headteachers put their “best” or most trusted staff to teach the pupils in the academic year they take place. For instance, with respect to the Year 6 SATs in England, Bradbury (2019) notes “*the pressure of SATs drives some of headteachers’ decision-making about staff allocation, encouraging them to prioritise Year 6 for additional staff, ‘stronger’ teaching assistants, and their ‘best’ teachers*”. Yet being chosen to teach what is often deemed to be the most important class also comes with pressure – potentially leading to anxiety about needing to live up to the expectations of others.

Increased workload

Those teachers most affected by high-stakes testing – whose pupils are taking the assessments – are likely to suffer from an increased workload. This may particularly be the case in the build-up to when the assessments take place, when they are most focused on helping young people to prepare. For instance, with respect to SATs, around one-in-ten schools hold booster sessions during the Easter holidays (Teacher Tapp 2019b), with around half of teachers believing pupils to be excessively drilled (Teacher Tapp 2019c). Teachers may also increase their use of practise tests, which they must then mark, to ensure pupils are as ready for the high-stakes assessment as possible. Previous work has shown that teachers working longer hours tend to have lower levels of wellbeing in the workplace (Jerrim and Sims 2021), particularly when extra time is devoted to non-teaching tasks.

Perceptions of reliability

Many of the issues discussed above could be worsened if teachers do not believe that the outcomes from high-stakes tests are reliable – regardless of whether this is true or not. It is likely to lead teachers to believe they have only limited control over the outcomes on which they (and their school) are being judged, which then leads them to feel anxious or stressed. Unfortunately, many primary school teachers in England feel this way about the Year 6 SATs; despite its good psychometric properties (Standards and Testing Agency 2018) around half of primary teachers do not believe SATs to be a reliable measure of pupil attainment (Teacher

Tapp 2022). Such feelings of limited control are likely to be exacerbated by the fact that teachers and schools are just one input affecting the results; pupil intake, parents and the home environment are vital as well (Leckie et al 2010). Together, such doubts about how much they can meaningfully influence the key metrics to which they will be held accountable may lead teachers to become more anxious about the tests.

Perceptions of cheating

When tests are high stakes, there is greater incentive to bend the rules. For those who engage in such unethical behaviour, their anxiety levels may increase due to either the fear of being caught or from the guilt of manipulating the results. Yet even for teachers that don't engage in such practises, their wellbeing may be impacted if they perceive cheating by others to be widespread – that the odds of success are stacked against them. Unfortunately, around one-in-five of those involved in the administration of Year 6 SATs say they have been encouraged to engage in questionable practises, such as pointing out incorrect answers to pupils (Teacher Tapp 2022).

Reacting to the emotional responses of others

Teachers are not the only group whose wellbeing may be affected by high-stakes tests – pupils may as well. Although previous quantitative analyses have suggested the aggregate effect is likely to be small (Jerrim 2021), polls have suggested that around 70% of Year 6 teachers reported at least some of their pupils to be upset or distressed about SATs (Teacher Tapp 2022). This may in turn impact the emotions of teachers. Indeed, previous research has suggested that such “emotional contagion” occurs in schools (Oberle and Schonert-Reichl 2016), leading to a feedback loop in the anxiety felt about high-stakes tests between teachers and pupils.

3. Data and methodology

Comparison groups

Given the arguments presented in section 2, if one wants to know more about how high-stakes assessments (such as the Year 6 SATs) impacts teacher wellbeing, what groups should be compared? Those leading Year 6 classes – whose pupils sit the SATs – are likely to be amongst the most affected. They will be tasked with preparing pupils for the tests and then held accountable for the results. Moreover, if schools’ drill pupils for SATs – as is often claimed – then this is likely to impact the workload of Year 6 teachers the most. Likewise, if SATs lead

to a narrowing of the curriculum, then this is likely to be most apparent in Year 6. They will also be working with pupils taking the SATs on a day-to-day basis, and thus be the staff who are most exposed to their emotions if they show any signs of distress. This is backed up by the qualitative research of Bradbury (2019), who notes that Year 6 teachers face “different pressures” to those teaching other year groups.

Who, then, should Year 6 teachers be compared to? The most logical comparators – similar in spirit to the approach of Gonzalez et al (2017) - are teachers less exposed to the pressures of the SATs (at least directly). Our primary comparison group is therefore primary school teachers leading Year 3 or Year 5 classes. Children in these year groups do not sit the SATs tests, nor any other national assessments in that academic year¹. Indeed, even More Than a Score (an anti-assessment interest group) notes that “*there are no high-pressure tests in Year 3/Year 5*” (More Than a Score 2022). Year 3 and Year 5 teachers will thus clearly be less affected by high-stakes assessments – and the SATs in particular – than Year 6 teachers. As a comparator, Year 3 and Year 5 teachers also have the advantage of working in primary schools, and thus otherwise have similar working environments as their Year 6 counterparts.

To test the robustness of our findings, we broaden our comparator group to also include primary teachers providing instruction to other school year groups. First, we add Year 4 teachers into our comparison group; although pupils in this academic year take a “multiplication tables check”, this test takes just five minutes and is a lot lower stakes than the SATs (Department for Education 2022). Second, we draw comparisons between Year 6 teachers and primary teachers working with all other school years. This has the advantage of maximising the sample size of our comparator group, though now includes teachers working with pupils of very different ages (between ages 4 and 10).

Data

The data we use were collected via the Teacher Tapp survey app (<https://teachertapp.co.uk/>). Each day at 1530, teachers receive a notification asking three short questions. On most days, around 5,000-8,000 teachers respond, and thus provides longitudinal information over time.

¹ Year 1 students have a phonics screening test, Year 2 pupils the Key Stage 1 SATs and Year 4 pupils the multiplication tables check. These are all much lower stakes assessments for teachers and schools than the Year 6 SATs. However, to get the cleanest possible comparator group, we focus on teachers of Year 3 and Year 5 pupils when no national assessments take place.

Although demographic characteristics of the Teacher Tapp panel is broadly in line with the population of teachers in England (see Teacher Tapp 2020 for further details), participants are self-selecting – and hence best considered a convenience sample. Survey weights are provided to ensure the sample matches the population on observable characteristics of school type, region, job role, gender and age.

At 16 points during the 2021/22 academic year, the Teacher Tapp panel were asked about their work-related anxiety. This was based on the UK Office for National Statistics general anxiety question (ONS 2015), which has been subject to extensive validation and is widely used (ONS 2012; Benson et al 2019):

“On a scale where 0 is ‘not at all anxious’ and 10 is ‘completely anxious’, overall, how anxious did you feel about work today?”

We have added “about work” to the standard question used by the ONS to ensure that teachers focus solely on their anxiety about work. Teachers were consistently asked this question on Tuesday afternoons to avoid possible day-of-week effects. We did not ask directly about SATs so (a) to not lead teachers into certain responses; (b) to ensure the question was also clearly relevant to our comparison groups and (c) to capture the impact on work-related anxiety overall. Appendix A provides further details on the precise dates this question was asked; importantly, this includes the week before, during and after the 2022 SATs took place². It also illustrates the distribution of responses, with the mean anxiety score at each timepoint sitting around 4.2 with standard deviation around 2.8.

Our main analysis includes teachers who responded to the work-related anxiety question on at least eight of the 16 occasions it was asked during the 2021/22 academic year. This includes up to 1,500 teachers, with specific details regarding sample sizes reported in Table 1³. In Appendix B we test the robustness of our key results to using more restrictive sample selection criteria (including only those teachers who responded to the work-related anxiety question on at least 13 occasions).

² As the SATs took place between 9th – 12th May 2022, we ensured our work-related anxiety question was asked on May 3rd, 10th and 17th to capture variation just before, during and after the test week.

³ Teachers who work in independent schools have been excluded from the analysis, due to their very different work environments to state schoolteachers (including many not administering SATs).

Descriptive information about the sample is presented in Table 1. This includes information previously collected about Teacher Tapp respondents such as demographic characteristics (e.g. age, gender, children at home, job role) and about the school in which they work (e.g. percent of disadvantaged pupils, Ofsted inspection rating). Figures are presented separately for Year 6 teachers and each comparison group. There are some clear differences in terms of experience, job role (e.g. whether the teacher holds a leadership position) and whether the teacher provides instruction across multiple year groups. This is consistent with the qualitative work of Bradbury (2019); certain types of teachers – most notably those who are more senior and experienced – are more likely to be assigned to teach Year 6 classes.

<< **Table 1** >>

Further descriptive information about the attitudes and behaviours of our comparison groups with respect to the SATs can be found in Table 2. This presents selected results from questions the Teacher Tapp panel has been asked about SATs since 2018⁴. It demonstrates how Year 6 teachers are more likely to run SATs booster sessions than primary teachers leading other year groups and that they are almost the most likely to have seen or been involved in maladministration of the tests (at least according to the most recent responses in 2022). Interestingly, on the whole, Year 6 teachers are somewhat less likely to want the SATs to be scrapped than teachers leading other school year groups. On the other hand, there is no clear difference across groups in perceptions of the reliability of the SATs or whether they are used as part of performance related pay.

<< **Table 2** >>

Methodology

Our analysis begins by plotting average work-related anxiety scores for our key comparator groups (e.g. Year 6 versus Years 3/5 teachers) at each survey date, with particular interest paid to variation around the SATs test week (9th – 12th May 2022)⁵. These results will provide a first descriptive account of how work-related anxiety compares across groups of teachers with different exposures to the SATs.

⁴ As different teachers have entered and exited the panel over time, the number and composition of the sample differs across the different timepoints.

⁵ Appendix H illustrates how the response rate to the work-related anxiety question is similar during the SATs test week as to most other points during the academic year.

These initial findings will then be formalised via regression modelling. Specifically, we will estimate OLS regression models of the form:

$$Anx_i = \alpha + \beta.Yr_6_i + \gamma.D_i + \delta.S_i + \varepsilon_i \quad (1)$$

Where:

Anx_i = Scores on the 0-10 work-related anxiety scale, averaged over the period under investigation (see below for further details).

Yr_6_i = A dummy variable capturing whether the teacher provides instruction to Year 6 pupils. Separate models will be estimated using Year 3/5, Year 3/4/5 and all other primary teachers as the reference group.

D_i = A vector of controls for teacher background characteristics. This includes age, gender, experience, job role, children at home and whether they teach across multiple year groups.

S_i = A vector of controls for school background characteristics. This includes Ofsted inspection rating and the percentage of pupils eligible for Free School Meals.

ε_i = Error term. Adjustments are made to the estimated standard errors to account for potential heteroskedasticity.

The parameter of interest is β . This captures the difference in average work-related anxiety between teachers most affected by the high-stakes SATs (Year 6 teachers) and those that are likely to be affected much less (the reference group – e.g. Year 3/5 teachers).

The model will be estimated several times, with the outcome (Anx_t) measured over different periods. To begin, for each teacher, we take the average anxiety score across each time they answered the question during the 2021/22 academic year (up until pupils take the SATs). The β estimate will thus capture whether Year 6 teachers generally feel more anxious about work throughout the academic year than our comparator groups. We then subsequently narrow the dates used to those closer to the SATs test week. For instance, we will re-estimate model (1) but with our work-related anxiety measure only being taken across the mid-February to mid-May period – i.e. capturing the immediate build up to and during the SATs test week. This is of course likely to be one of the busiest and most stressful periods for Year 6 teachers as they hurriedly prepare pupils for the assessment. Likewise, we will also present separate estimates using only responses to the anxiety question in the week before, during and after the SATs take place. Together, these estimates reveal whether teachers who are most exposed to the pressure

of SATs differ in their anxiety levels to those groups with less exposure, and when in the academic year any difference is most acute.

Robustness tests

Several robustness tests will be conducted – reported in the online supplementary material – to explore the sensitivity of the results. First, we will exclude headteachers and senior leadership teams members from the sample, thus focusing on how the SATs are related to the work-related anxiety of class teachers and middle leaders only. Second, we will exclude from the sample any teacher who teaches across multiple year groups. This will allow us to focus on teachers who only teach Year 6 pupils, and how they compare to teachers who only teach (for instance) Year 3 or Year 5 pupils. Third, we will further restrict the sample so that it only includes those who responded to the work-related anxiety question at least 13 of the 16 times it was asked during the 2021/22 academic year (as compared to a minimum of 8 times in the main analysis). Finally, alternative descriptive results will be presented when we dichotomise the outcome variable, investigating how the SATs are associated with the percentage of teachers who report *high* anxiety levels (scores of 7 and above on the 0-10 anxiety scale)⁶.

Limitations / threats to validity

While we believe that our approach has some important advantages over previous work – most notably in thinking about and drawing explicit comparisons to meaningful “control” groups – we also recognise there remains some limitations. For instance, although the 2021/22 academic year had largely returned to normal, schools still faced some disruption between December 2021 and mid-February 2022 from the Omicron COVID wave. Although schools remained open, there were quite high levels of staff and pupil absence during this period. Despite our comparison groups probably being affected in similar ways to Year 6 teachers, one cannot completely discount this as having some impact on our results. Importantly, though, the period leading up to the SATs – from around mid-February onwards – schools had largely returned to normal.

Perhaps more important was the lasting impact that COVID had on how the results from the SATs were to be used. In July 2021 – almost a year before the tests took place – the Department for Education in England announced that it would not publish SATs performance tables in

⁶ This follows the cut-off for high anxiety levels used by the UK’s Office for National Statistics (ONS) using this scale.

2022. This means that one of the key factors potentially make SATs stressful – the open publication of results – did not take place. The results, however, were still accessible to school leaders, managers, governors and Ofsted (the school inspectorate). Nevertheless, in 2022, SATs were arguably slightly lower stakes than in a “normal” year.

Finally, as noted above, teachers are not randomly assigned to year groups. Rather, senior leaders allocate staff based on their attributes and skills. It could be that headteachers choose to not ask staff to teach Year 6 if they are prone to anxiety issues or do not think they can cope with the pressure surrounding the SATs. The controls included in our regression models are only likely to partially address the potential confounding effect of teachers with different characteristics being deployed to teach different year groups.

4. Results

Figure 1 compares mean work-related anxiety scores for Year 6 teachers (black solid line with square markers) to Year 3/5 teachers (dashed grey line with circular markers) at 16 points during the 2021/22 academic year. There are three key points to note.

<< Figure 1 >>

First, in general, Year 6 teachers had slightly *lower* levels of work-related anxiety than Year 3/5 teachers at most timepoints. In other words, teachers who were most exposed to the pressures of SATs were slightly *less* anxious about work than their peers with less exposure. The difference stands at around ½ of one point on the 0-10 anxiety scale, with a mean score of approximately 4 for Year 6 teachers and 4.5 for Year 3/5 teachers. This is equivalent to an effect size of around 0.2 of standard deviations. Caution is needed when interpreting these unconditional results, however, because they could be driven by selection; as Table 1 illustrates, more senior and experienced teachers are more likely to be assigned to teach Year 6 pupils. Indeed, one plausible explanation for the generally lower levels of anxiety of Year 6 teachers in Figure 1 is that headteachers are assigning less anxious staff to teach this year group.

Second, the only point where the opposite hold true – i.e. where Year 6 teachers have higher levels of anxiety than our Year 3/5 comparison group – is during the week the SATs take place. The magnitude of the difference is similar to before - around ½ an anxiety point (~0.2 standard deviations) – but now in the opposite direction. This provides the first suggestion that the SATs negatively impacts the wellbeing of Year 6 teachers, but only during the narrow window when the tests take place.

Finally, the week after the SATs have finished – and through to the end of the academic year – the status quo returns. The work-related anxiety of Year 6 teachers has quickly returned to “normal” levels, being slightly lower than the Year 3/5 comparison group. This in-turn suggests that the negative impact of the SATs on the anxiety levels of Year 6 teachers is relatively short-lived; there is a short, sharp shock when the tests take place, but this is then quickly reversed.

Table 3 formalises these results by presenting estimates from our regression models. Figures capture the difference in average anxiety scores between Year 6 teachers and each comparator group, conditional on the background characteristics controlled. The top row refers to differences in average anxiety levels across groups throughout the 2021/22 academic year (up until the week of May 10th when the SATs took place), while the second row takes average anxiety levels during the three-month period building up to the test week. The bottom three rows present analogous estimates of teacher anxiety the week before, during and after the SATs took place.

<< Table 3 >>

On average, there is no evidence that Year 6 teachers felt more anxious about work throughout the academic year than any of the three comparator groups. Although the point estimates are negative (suggesting, as per Figure 1, Year 6 teachers might have slightly lower anxiety levels than Year 3/5 teachers) these do not reach statistical significance at the five percent level. The same holds true with respect to the three-month period leading up to the SATs. Thus, overall, there is no evidence that those teachers most exposed to the pressures of the SATs are generally more anxious about work than their peers with less exposure during most of the academic year, including in the three-month period leading up to the tests.

Turning to the bottom three rows – focusing on the week immediately before, during and following the SATs – a clear result emerges. There is no clear difference in work-related anxiety of Year 6 teachers and any of our comparator groups in the week prior and following the SATs. However, during the SATs week, a substantial difference is found. The anxiety of Year 6 teachers is around 0.7-0.8 anxiety score points above those of our comparator groups, which is equivalent to around 0.25 standard deviations. These are the only estimates that are statistically significant at the five percent level. This is consistent with the descriptive results presented in Figure 1.

To conclude, Figure 2 presents estimates of the *conditional* difference in anxiety levels between Year 6 and Year 3/5 teachers at each survey date. Positive values indicate where Year 6 teachers have higher conditional average anxiety levels, with solid markers indicating where the difference is statistically significant at the five percent level. Overall, the story from these results is similar to before. Through most of the academic year, Year 6 teachers anxiety levels are around 0.3 points (on the 0-10 scale) below that of Year 3/5 teachers, while during the week of the SATs they are 0.7 anxiety points above. This represents an increase in work-related anxiety during the test week of around 0.35 standard deviations, which quickly reverses once the SATs are complete. Together, this provides evidence that the SATs have a short, sharp shock on teachers' anxiety levels during the test week of the test, but little sign of there being a prolonged effect.

<< **Figure 2** >>

Heterogeneity across sub-groups

Table 4 repeats the analysis presented in Table 3, but now with separate estimates presented for six sub-groups: (a) male teachers; (b) female teachers; (c) teachers who believe the SATs are reliable; (d) teachers who believe SATs are unreliable; (e) teachers who work in schools holding SATs booster lessons and (f) teachers who work in schools that do not hold SATs booster lessons. Note that in this table we focus on comparisons between Year 6 and Year 3/4/5 teachers to ensure these sub-group estimates are based on a sufficient sample size.

<< **Table 4** >>

Overall, the pattern of results is very similar across each of these sub-groups. In each case, there is no difference in anxiety levels between Year 6 and Year 3/4/5 teachers, except during the SATs test week. This includes in the weeks and months leading up to when the SATs take place. Thus, overall, it seems that our key finding – that the SATs increase anxiety levels but only during the test week – holds across each of the sub-groups considered.

Sensitivity analyses

Appendix B presents alternative estimates where the sample is restricted to teachers who responded to the work-related anxiety question on at least 13 of the 16 occasions (rather than 8 occasions in the main analysis above). Results are very similar to those presented in Figure 2. Appendix C and D provide alternative versions of Figures 1 and 2 but now drawing

comparisons between Year 6 teachers and (a) Year 3/4/5 teachers (b) all other primary teachers. These demonstrate that our key results are robust to whichever comparison group is used. When headteachers and senior leaders are removed from the analytic sample in Appendix E, an even sharper peak in work-related anxiety is observed during the SATs test week. Otherwise, the patterns observed remain unchanged. Likewise, our findings remain intact when the sample is restricted to only those teachers who provide instruction to a single school year group (see Appendix F). Finally, Appendix G provides alternative descriptive results focusing on the percent of teachers reporting high anxiety levels (rather than looking at mean scores). This demonstrates how the percent of Year 6 teachers with high levels of work-related anxiety increases from around 25% during most of the academic year to around 35% during the week of the SATs.

5. Conclusions

High-stakes assessments are an important part of education accountability systems around the world. Yet their use is not without controversy, particularly in primary schools. Proponents of high-stakes tests argue that they provide the information needed to hold schools, managers and education leaders to account, raising standards as a result (Supovitz 2009). Their detractors however point towards the many unintended negative consequences they believe more than offsets any of the potential gains. One such claim is that high-stakes assessments negatively impact mental health and wellbeing in the education sector (Hogberg and Horn 2022). This not only includes those pupils taking the tests, but the staff that teach them as well (Gonzalez et al 2017).

Yet there are some key issues with existing evidence in this area. Much of the previous work is small scale, cross-sectional and only captures opinions about such tests from school staff. Few have attempted to measure mental health objectively using large-scale longitudinal data, or attempted to look at differences in wellbeing levels across meaningful comparison groups (e.g. those teachers more versus less impacted by the tests). This paper has started to make inroads into this gap in the evidence base, presenting new empirical insights into the link between the Year 6 SATs and the anxiety levels of teachers in England.

Our results provide some evidence that high-stakes assessments such as SATs are related to how anxious teachers feel about work. In particular, there seems to be a short, sharp shock to teachers' anxiety levels during the week when the tests take place. At the same time, there is little evidence of a substantial increase in work-related anxiety in the period leading up to the

tests. Hence the negative impact of SATs on teachers' anxiety levels seems to be over relatively quickly. This could be due to either the SATs not inherently being stressful to teachers - outside of the test week - or to headteachers managing the pressures through how they allocate staff. Either way, the result seems to be that the "impact" of SATs on teacher wellbeing is probably quite limited.

We recognise our work also has its limitations. First, like most previous studies in this area, participants are self-selected and are not drawn from a national probability sample. Future work using data from such a sample design – and achieving a high response rate – would further increase one's confidence in the generalisability of the results. Second, we have focused on one specific dimension of teachers' wellbeing at work (anxiety). Although this is one of the dimensions of wellbeing most likely to be impacted by high-stakes assessments, one cannot rule out the association being different for other areas (e.g. happiness, life-satisfaction). Third, following the COVID-19 pandemic, the 2022 SATs were not completely back to "normal". Although many of the key stressors associated with such tests were back in place, at least one key element – the publication of school league tables – was not. While we do not believe this is likely to have a major impact on our substantive findings, further data collection covering another academic year would be needed for this to be empirically verified. Finally, as noted in the methodology section, headteachers may select certain types of teachers (e.g. their most capable, ambitious, experienced or cool-headed) to teach Year 6 pupils. This may in turn confound some of our results. Thus, like other existing studies into the link between high-stakes assessments and teacher wellbeing, our estimates capture conditional associations rather than establishing cause and effect. Further work – both qualitative and quantitative - is needed into understanding who school leaders deploy to teach those pupils who are about to take high-stakes tests.

What might our findings nevertheless imply for the future of high-stakes assessments in primary schools? On the one hand, one may argue that we have shown there to be a clear link between the presence of such tests and teachers' anxiety levels, even if the negative effects appear relatively short-lived. On the other hand, one might argue that we all experience stressful periods at work, and that in this respect teachers are little different from other professionals. For accountants it is the end of the financial year, for those working in the emergency services it is Friday and Saturday nights and - as we have shown – for Year 6 teachers in England it is the week when the SATs take place. There are hence likely to be

stronger cases to be made against the use of high stakes assessments in primary schools than the limited link observed with their wellbeing.

References

Berliner, D. 2011. Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education* 41(3): 287-302.

Bradbury, A. 2019. Pressure, anxiety and collateral damage. The headteachers' verdict on SATs. More Than A Score. Accessed 24/08/2022 from <https://www.morethanascore.org.uk/wp-content/uploads/2019/09/SATs-research.pdf>

Brockmeier, L.; Green, R.; Pate, J.; Tsemunhu1, R. and Bochenko, M. 2014. Teachers' beliefs about the effects of high stakes testing. *Journal of Education and Human Development* 3(4): 91-104.

Department for Education. 2022. Information for parents: 2022 multiplication tables check. Department for Education, England. Accessed 01/09/2022 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1031901/2022_Information_for_parents_Multiplication_tables_check.pdf

Gonzalez, A.; Peters, M.; Orange, A. and Grigsby, B. 2017. The influence of high-stakes testing on teacher self-efficacy and job-related stress. *Cambridge Journal of Education* 47(4): 513-531, DOI: [10.1080/0305764X.2016.1214237](https://doi.org/10.1080/0305764X.2016.1214237)

Demir, C. and Keleş, O. 2021. The impact of high-stakes testing on the teaching and learning processes of mathematics. *Journal of Pedagogical Research* 5(2): 119-137

Gunn, J.; Al-Bataineh, A.; Abu Al-Rub, W. 2016. Teachers' perceptions of high-stakes testing. *International Journal of Teaching and Education* 4(2): 49-62.

Hargreaves, A. 2020. Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change* 21: 393-420.

Högberg, B. and Horn, D. 2022. National high-stakes testing, gender, and school stress in Europe: A difference-in-differences analysis. *European Sociological Review* <https://doi.org/10.1093/esr/jcac009>

Jerrim, J. 2021. National tests and the wellbeing of primary school pupils. *Assessment in Education: Principles, Policy & Practice* 28(5/6): 507-544.

Jerrim, J. and Sims, S. 2021. When is high workload bad for teacher wellbeing? Accounting for the non-linear contribution of specific teaching tasks. *Teaching and Teacher Education* 105. <https://doi.org/10.1016/j.tate.2021.103395>

Leckie, G.; Pillinger, R.; Jenkins, J. and Rasbash, J. 2010. School, family, neighbourhood: which is most important to a child's education? *Significance* 7: 67-70. <https://doi.org/10.1111/j.1740-9713.2010.00421.x>

Lotz, C. and Sparfeldt, J. 2017. Does test anxiety increase as the exam draws near? – Students' state test anxiety recorded over the course of one semester. *Personality and Individual Differences* 104: 397-400.

Mathison, S. and Freeman, M. 2006. Teacher stress and high stakes testing. How Using One Measure of Academic Success Leads to Multiple Teacher Stressors. Pp 43-63 in *Understanding Teacher Stress in an Age of Accountability*. Accessed 24/08/2022 from https://d1wqtxts1xzle7.cloudfront.net/34533890/003-Stress_Mathison_Freeman-with-cover-page-v2.pdf?Expires=1661349710&Signature=g8Lp8QquTpALcmffIM5j0cA4OjUeMIjhSOOk~eBgPHN5EdwQpZnhFICe5xIUPBnyQk9hMG0YBYSIGr6bEcUDZRYK6nM-2WAYkZRYjjLE1WNRx6OWEJxzfYCUgKeMIN3hpdR0ioELIDKu1Bsn1lVhT1qy~MhGInkgFOSFaG3gZq~dr6Szvqfsg-d7U2ew9A7oyWgr3tu9kP6yYVo8xWjMaENpn~3Wqp6s8gF-pQKpcvmSf8PtbB3-R3xpMqB96-1LopoK4BaXUGc24YJp1c6SqsZXV38WpastCDd1FdYRJqadIXiVZlwagv1c-AjxWiZ8TvEp5WPCh9eqGX33OstyQ_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

More Than a Score. 2022. A parents' guide to primary testing. Accessed 26/08/2022 from <https://www.morethanascore.org.uk/what-we-do/parents-guide/year-5/?scrollto=content>

National Education Union. 2019. Too much testing. Accessed 24/08/2022 from [NEU1303 Scrap the SATs ballot crib sheet for web v1_0.pdf](#)

Neumann, E.; Towers, E.; Gewirtz, S. and Maguire, M. 2016. A curriculum for all? The effects of recent Key Stage 4 curriculum, assessment and accountability reforms on English secondary education. Kings College London, London. Accessed 01/09/2022 from <https://www.kcl.ac.uk/archive/news/education/news/curriculum-for-all-final.pdf>

Ofsted. 2019. Exploring the issue of off-rolling. Accessed 24/08/2022 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/936524/Ofsted_offrolling_report_YouGov_090519.pdf

Oberle, E. and Schonert-Reichl, K. 2016. Stress contagion in the classroom? The link between classroom teacher burnout and morning cortisol in elementary school students. *Social Science & Medicine*, 159: 30-37.

Page, D. 2017. Conceptualising the surveillance of teachers. *British Journal of Sociology of Education* 38(7): 991-1006.

Perryman, J. and Calvert, G. 2020. What motivates people to teach, and why do they leave? Accountability, performativity and teacher retention. *British Journal of Educational Studies* 68(1): 3-23.

Reese, M. Gordon, S. and Price, L. 2004 Teachers' perceptions of high-stakes testing. *Journal of School Leadership* 14(5):464-496. doi:[10.1177/105268460401400501](https://doi.org/10.1177/105268460401400501)

Saminsky, A. 2011. High-stakes standardized testing: A panacea or a pest? *Inquiries Journal/Student Pulse* 3(01).

Smith, W. and Holloway, J. 2020. School testing culture and teacher satisfaction. *Educational Assessment, Evaluation and Accountability* 32: 461–479.

Standards and Testing Agency. 2018. National curriculum test handbook: 2018 Key stages 1 and 2. Standards and Testing Agency, London, England. Accessed 01/09/2022 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/765749/2018_NCT_Handbook_PDFA.pdf

Supovitz, J. 2009. Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change* 10: 211–227

Teacher Tapp. 2019a. Election pledges, performance-related pay and three other remarkable findings from this week. Accessed 25/08/2022 from <https://teachertapp.co.uk/manifesto-pledges-performance-related-pay/>

Teacher Tapp. 2019b. Genies, GCSEs, and the 11% of primary schools doing SATs booster classes. Accessed 01/09/2022 from <https://teachertapp.co.uk/genies-gcse-sats-booster-revision/>

Teacher Tapp. 2019c. SATs reform: 50 shades of playing the system (and how to reform it). Accessed 01/09/2022 from <https://teachertapp.co.uk/sats-reform-change-future-teacher-assessment-reliable>

Teacher Tapp. 2020. Who is on Teacher Tapp? (And why?!). Accessed 15/08/2022 from <https://teachertapp.co.uk/not-like-others-teacher-tapp-who-sample-representative-reweights/>

Teacher Tapp. 2022. All things SATs and what's going on with study leave? This, and more... Accessed 25/08/2022 from <https://teachertapp.co.uk/all-things-sats-and-whats-going-on-with-study-leave-this-and-more/>

TES. 2018. Exclusive: Sats take toll on teachers' mental health. Accessed 24/08/2022 from <https://www.tes.com/magazine/archive/exclusive-sats-take-toll-teachers-mental-health>

Tucker, F. and Horton, J. 2019 The show must go on! Fieldwork, mental health and wellbeing in Geography, Earth and Environmental Sciences. *Area* 51: 84-93.

West, A. and Pennell, H. 2006. Market-oriented reforms and "high stakes" testing: incentives and consequences In: Vinokur, Annie, (ed.) *Pouvoirs et Mesure En Éducation. Cahiers de la recherche sur l'education and savoirs* (1/2005). Maison Des Sciences De L'homme, Paris, France, 181-199. ISBN 9782735110803.

Table 1. Descriptive statistics for the analytic sample

	Comparison group			Year 6
	Year 3/5	Year 3/4/5	All	
Ofsted rating				
Outstanding	15	14	15	17
Good	74	75	73	69
Requires Improvement	11	11	12	14
FSM Quartile				
Q1 (least deprived)	19	20	22	22
Q2	26	27	28	30
Q3	29	27	26	22
Q4 (most deprived)	27	27	24	26
Gender				
Male	20	21	16	23
Female	80	79	84	77
Job role				
Class teacher	51	52	49	26
Middle leader	28	28	27	33
Senior Leadership Team	19	17	20	33
Headteacher	2	2	3	7
Children				
None	40	42	41	40
Under 5	14	14	11	13
Primary age	17	14	16	16
Secondary age	9	9	10	8
Over 16	11	12	12	13
Missing data	9	9	10	10
Age				
20s	24	25	23	15
30s	31	31	32	39
40s	35	32	33	33
50+	11	12	12	14
Experience				
Under 5 years	29	27	26	16
5 to 10 years	23	26	24	27
10 to 20 years	28	28	31	36
Over 20 years	20	19	20	21
Teach multiple year groups				
No	82	78	79	56
Yes	18	23	21	44
N	327	531	887	613

Notes: Figures refer to column percentages.

Table 2. Differences in attitudes and behaviours towards the SATs across comparison groups

	N (all primary)	Comparison group		All	Year 6
		Year 3/5	Year 3/4/5		
1. % Think that SATs are reliable					
18/01/2018	344	40%	41%	46%	45%
11/05/2022	1,740	46%	47%	46%	46%
2. % Seen / done maladministration					
03/07/2018	459	29%	30%	25%	27%
23/05/2019	903	15%	18%	18%	19%
13/05/2022	687	21%	20%	28%	53%
3. % Performance related pay linked to results					
24/08/2018	688	78%	73%	74%	75%
06/11/2018	651	63%	66%	69%	72%
21/11/2019	2,169	60%	60%	62%	64%
10/11/2021	2,553	45%	44%	43%	41%
4. % Running a booster session over Easter					
05/03/2019	951	8%	8%	10%	18%
08/03/2020	1,717	5%	5%	8%	14%
5. % That would scrap SATs					
17/04/2019	1,300	68%	70%	69%	60%
03/12/2020	809	73%	71%	69%	62%
22/06/2022	1,813	46%	47%	43%	48%

Notes: Questions include variations of 1. “How reliable do you think the year 6 SATs are as a measure of pupil attainment?”, 2. “If you have been directly involved in administering Year 6 SATs - either in your current school or in an earlier school - have you ever done or been encouraged to do any of the following practices?”; 3. “Does one of your performance related pay management targets refer to a specific percentage of exam/SATs results?”; 4. “Will you personally be running any exam (e.g. SATs or GCSE) booster/revision classes over the Easter holidays?” 5. “The government is scrapping KS1 SATs but is keeping most KS2 SATs. What would you do next year if you were in charge?”. Note that the question asked for (5) differs across time points, and thus the figures are not directly comparable. Sample of respondents varies across the time points that the questions were asked.

Table 3. OLS regression model estimates of differences in work-related anxiety between Year 6 teachers and selected comparison groups

	Years 6 vs 3/5		Years 6 vs 3/4/5		Years 6 vs all other	
	Difference	SE	Difference	SE	Difference	SE
2021/22 average	-0.32	0.17	-0.23	0.15	-0.10	0.14
mid-Feb to mid-May	-0.31	0.18	-0.21	0.16	-0.09	0.15
Week before SATs	0.16	0.24	0.08	0.21	0.22	0.20
SATs test week	0.68*	0.25	0.77*	0.22	0.83*	0.20
Week following SATs	-0.21	0.25	-0.09	0.22	-0.04	0.20

Notes: Figures refer to the difference in average anxiety scores between Year 6 teachers and the stated comparison group. Positive figures indicate where Year 6 teachers report higher levels of work-related anxiety than the comparator. The first row (Year 6 average) uses the average anxiety of teachers between the start of the school year and the SATs test week. The second row (mid-Feb to mid-May) averages teachers work-related anxiety scores following the February half-term until the SATs test week. The final three rows report differences in anxiety in the week before, during and after when the SATs take place. * indicates statistical significance at the five percent level.

Table 4. OLS regression model estimates of differences in work-related anxiety between Year 6 and Year 3/4/5 teachers. Estimates for sub-groups.

	Male	Female	Believe SATs unreliable	Believe SATs reliable	Booster lessons for Year 6	No booster lessons for Year 6
2021/22 average	-0.39	-0.22	-0.08	-0.24	-0.28	-0.18
mid-Feb to mid-May	-0.49	-0.17	0.08	-0.23	-0.29	-0.11
Week before SATs	-0.50	0.16	0.21	-0.16	-0.05	-0.17
SATs test week	0.68	0.76*	0.93*	0.69*	0.73*	0.85*
Week following SATs	0.12	-0.14	-0.07	-0.18	-0.12	-0.09

Notes: Figures refer to the difference in average anxiety scores between Year 6 teachers and Year 3/4/5 teachers as the comparison group. Positive figures indicate where Year 6 teachers report higher levels of work-related anxiety. The first row (Year 6 average) uses the average anxiety of teachers between the start of the school year and the SATs test week. The second row (mid-Feb to mid-May) averages teachers work-related anxiety scores following the February half-term until the SATs test week. The final three rows report differences in anxiety in the week before, during and after when the SATs take place. * indicates that the difference between Year 6 and Year 3/4/5 teachers is statistically significant at the five percent level.

Figure 1. Average work-related anxiety scores of Year 6 and Year 3/5 teachers during the 2021/22 academic year

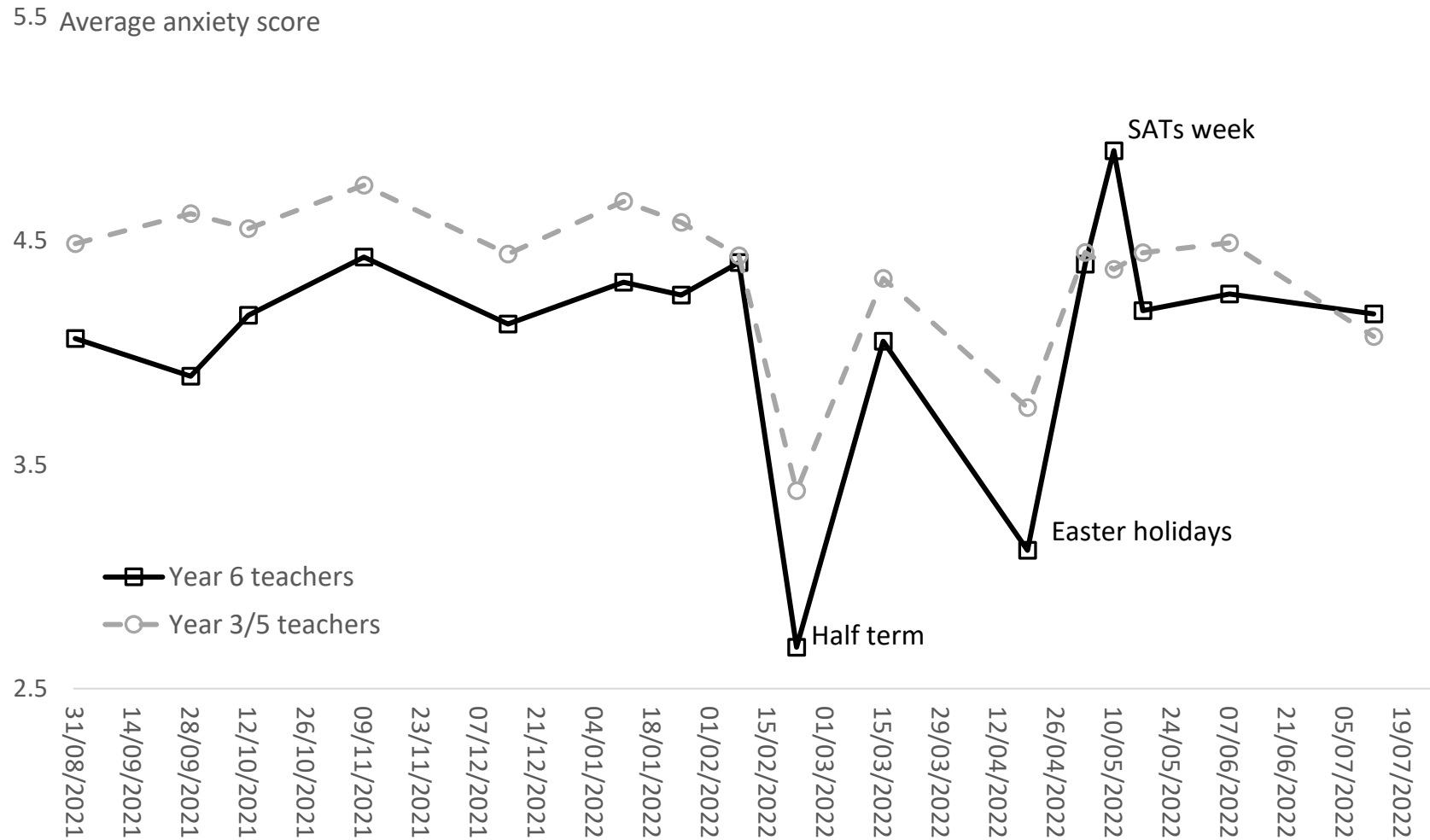
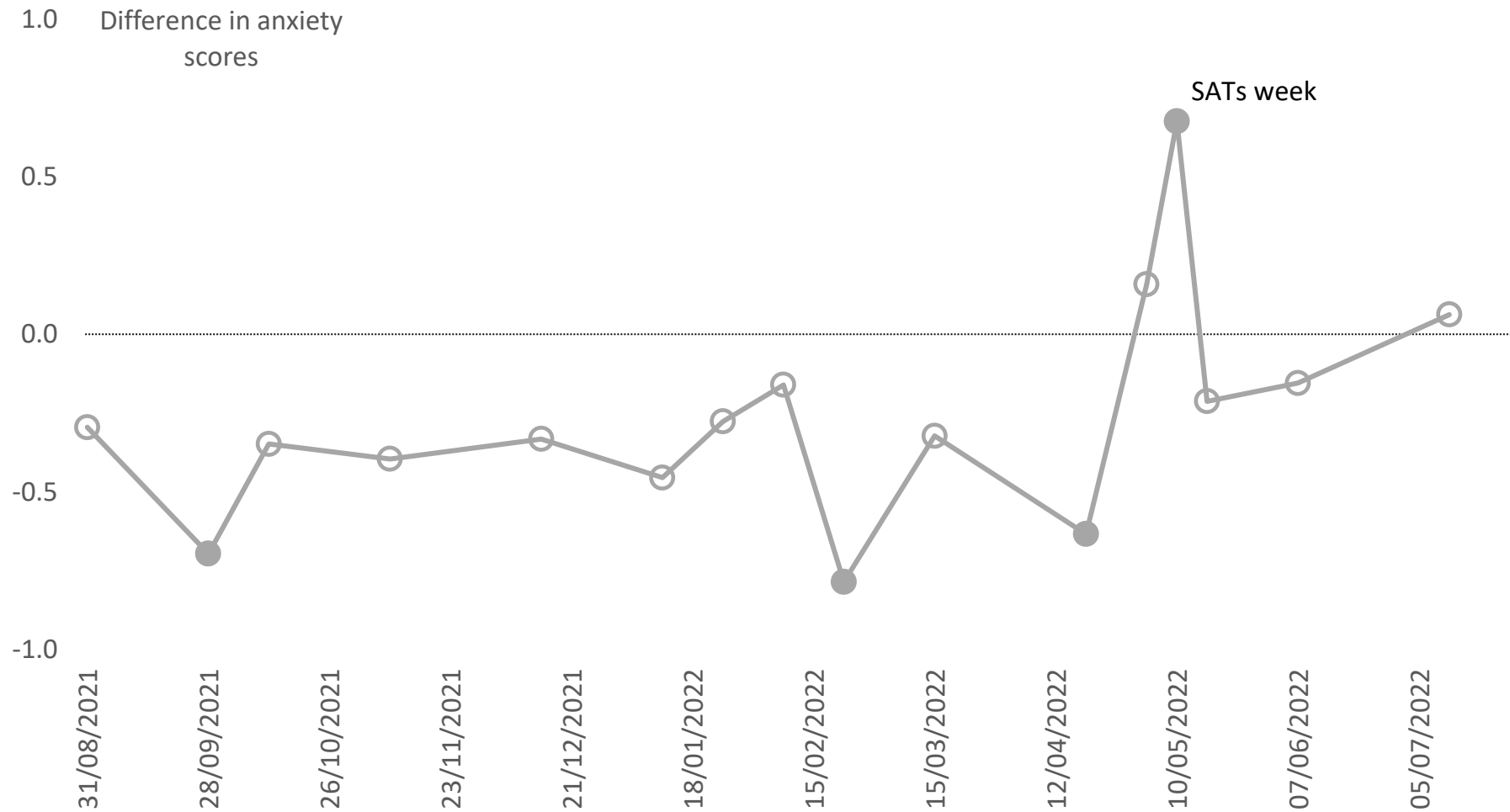


Figure 2. Difference in *conditional* work-related anxiety scores of Year 6 and Year 3/5 teachers during the 2021/22 academic year



Notes: Figures refer to the difference in work-related anxiety between Year 6 teachers and Year 3/5 teachers. Positive figures are where Year 6 teachers report higher levels of work-related anxiety than Year 3/5 teachers. Circular markers with solid fill illustrate where the difference is statistically significant at the five percent level (hollow markers indicate where the difference is not statistically significant).