

Oxford Review of Education

ISSN: 0305-4985 (Print) 1465-3915 (Online) Journal homepage: www.tandfonline.com/journals/core20

Test anxiety: Is it associated with performance in high-stakes examinations?

John Jerrim

To cite this article: John Jerrim (2023) Test anxiety: Is it associated with performance in high-stakes examinations?, Oxford Review of Education, 49:3, 321-341, DOI: 10.1080/03054985.2022.2079616

To link to this article: https://doi.org/10.1080/03054985.2022.2079616

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 14 Jun 2022.

Submit your article to this journal

Article views: 39389

View related articles

View Crossmark data

Citing articles: 26 View citing articles

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=core20

OXFORD REVIEW OF EDUCATION
2023, VOL. 49, NO. 3, 321–341
https://doi.org/10.1080/03054985.2022.2079616

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Test anxiety: Is it associated with performance in high-stakes examinations?

John Jerrim

Department of Social Science, UCL Social Research Institute, London, United Kingdom of Great Britain and Northern Ireland

ABSTRACT

A long-established literature has found that anxiety about testing is negatively related to academic achievement. Yet there remains some debate as to whether this is simply due to less academically able pupils being more likely to develop education-related anxiety issues. This paper presents new evidence on this matter, focusing upon how test anxiety – as measured by five questions included in the PISA 2015 survey – is related to the grades 15/16-year-olds achieve in England's high-stakes GCSE examinations. I find little evidence that teenagers with low or high levels of test anxiety achieve lower GCSE grades than pupils with average levels of test anxiety. Thus, in contrast to much of the existing literature, no clear relationship between test anxiety and examination performance is found.

KEYWORDS

Test anxiety; high-stakes testing; PISA; examination performance

1. Introduction

Anxiety about testing and schoolwork may impact the grades that young people achieve in high-stakes examinations. This effect could either be positive or negative. For some, anxiety about an upcoming examination could lead to increased levels of motivation, focus, effort and subsequently higher grades (Kader, 2016). Yet, for others, the effects could be debilitating. It may, for instance, lead to worry and an inability to concentrate (and/or a tendency to procrastinate) in the weeks and months building up to the examination period, limiting their ability to work and revise material effectively (Cassady, 2004; Howard, 2020; Keogh et al., 2004; Putwain & von der Embse, 2018). Test anxiety may also lead to problems during the examination itself, leading to an inability to focus or forgetting key content, such as 'going blank' in the exam (Doctor & Altman, 1969). Given the importance of GCSEs in England, it is vital that a better understanding of this issue is developed.

This study therefore builds upon previous work that has investigated the link between test anxiety and examination performance. In a recent meta-analysis, Von der Embse et al. (2018) found 'a consistent pattern of relationships with higher levels of test anxiety and lower levels of performance, across various testing formats'. The authors also note,

CONTACT John Jerrim J.Jerrim@ucl.ac.uk Department of Social Science, UCL Institute of Education, University College London, 20 Bedford Way London, London WC1H 0AL, United Kingdom of Great Britain and Northern Ireland
Supplemental data for this article can be accessed online at https://doi.org/10.1080/03054985.2022.2079616

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

however, how there remains questions surrounding the direction of this relationship, and whether it captures cause and effect. Indeed, the Von der Embse et al. (2018) meta-analysis does not explicitly state whether selection bias has been accounted for, with no mention of this within the study's inclusion criteria. Their findings are however consistent with an earlier meta-analysis conducted by Hembree (1988), who concluded that 'test anxiety causes poor performance'.

The work of Cassady and Johnson (2002) found that high levels of test worry are associated with lower Scholastic Aptitude Test (SAT) scores, and thus 'support the conclusion that cognitive test anxiety exerts a significant stable and negative impact on academic performance measures'. In contrast, qualitative research conducted by Putwain (2009), based upon 34 semi-structured interviews in England, found both debilitating and facilitating effects of test anxiety. He noted how test anxiety led some students to be more 'motivated, out of a fear of failure, to adopt a more contentious approach to examination preparation and/or make a greater effort'. Potential differences in the effects of test anxiety upon academic performance for different sub-groups has also been investigated (Putwain, 2008). Although he found girls to be more test anxious on average than boys, there was no evidence that this differentially impacted their GCSE performance. Some differences were observed, however, by socio-economic status.

Yet Sommer and Arendasy (2014) argue that the negative relationship between test anxiety and examination performance is largely due to selection bias (i.e. less able pupils tend to be more anxious which, if unaccounted for, leads to an overestimation of the negative effect of anxiety upon test performance). This is supported by Howard (2020) who, after conducting a review of the literature, concludes that 'after controlling for ability, high levels of test anxiety are generally associated with small reductions in test performance'. Other work has reached similar conclusions too. Everson et al. (1989) investigate the relationship between test anxiety and standardised test performance, controlling for prior achievement. They find that the 'cognitive component of test anxiety (worry), and prior academic achievement contribute independently, not interactively, to performance'. Musch and Broder (1999) also look at the relationship between test anxiety and performance, conditioning upon a measure of prior achievement. They find that 'both [prior] maths skill and test anxiety added unique variance in explaining performance'. Sommer and Arendasy (2015) also investigated the link between test anxiety and examination performance in a high-stakes setting (medical school entrance exams). Using structural equation modelling and item response theory, their results are consistent with the 'deficit model', which suggests that test anxiety and performance are not causally linked. Using data from a sample of university students, Reeve and Bonaccio (2008) also found results consistent with the deficit model of test anxiety, noting that their results 'seem to suggest that high anxiety simply accompanied lower ability'.

Such studies have led to an impressive evidence base about the relationship between test anxiety and academic performance. Yet some important gaps remain. For instance, although large-scale meta-analyses have confirmed test anxiety and academic achievement are negatively correlated, it is less clear the extent to which this is due to selection bias, with academically weaker pupils displaying higher levels of anxiety (Von der Embse et al., 2018). Thus, under the so-called 'deficit model' (Tobias, 1990), the negative

correlation between test anxiety and future achievement is spurious. It is hence unclear how strong the anxiety-achievement relationship is once this issue has been taken into account. Moreover, despite clear theoretical reasons to believe that the relationship between test anxiety and academic achievement may be non-linear, there has been little empirical evidence exploring this issue. (One important exception is Sung et al. (2016), who found that for low-achieving pupils greater levels of test anxiety was found to be positively related to achievement, while for high-achieving pupils the relationship was negative). This could, in turn, mask some important academic and education policy issues. For instance, against conventional wisdom, it might be that some degree of test anxiety has benefits for young people, motivating them to prepare for their exams and to take revision (and other aspects of preparation) seriously. Similarly, at what point do high levels of test anxiety start to exert a substantial negative effect upon achievement? Does this affect the grades of a large proportion of the population, or is it confined to particularly large negative effects amongst the most anxious five or ten per cent of pupils?

This paper provides new evidence on these issues. Using Programme for International Student Assessment (PISA) data linked to national administrative records in England, the paper considers how test anxiety is related to the grades young people achieve in the high-stakes GCSE exams. The rich data available allows us to explore how the strength of this relationship changes under different assumptions about how prior achievement is controlled (which may be driving the selection effects leading to the negative correlation reported in most of the literature). The paper also explicitly considers whether the anxiety-achievement relationship is indeed linear, or if large negative effects for some students (e.g. those who are very anxious) are to some extent offset by positive effects for others (e.g. students with very low levels of anxiety who may be overly relaxed). In doing so, it provides important new evidence on how test anxiety is related to the grades young people achieve in a set of high-stakes examinations, in an empirical setting where analysis of large-scale, nationally representative longitudinal data addressing this issue has been relatively sparse.

2. Theoretical background and research questions

Test anxiety has been defined as '*the subjective experience of intense physiological, cognitive and/or behavioural symptoms of anxiety before or during test-taking situations that interferes with test performance*' (Sawka-Miller, 2011). It is often divided into two separate factors: emotionality and worry (Minor & Gold, 1985). The former refers to the immediate, physiological symptoms induced by sitting an exam, such as an increase in heart rate, feeling sick or panic. Emotionality responses typically occur during – or in very close proximity to – the test or examination that is causing the anxious response. Worry, on the other hand, '*refers to the cognitive component of test anxiety, such as negative and derogatory self-statements related to failure*' (Putwain, 2007). In other words, it is the negative thoughts young people have about evaluation of their performance, feelings of unpreparedness for the exam and the negative consequences of failure. Such feelings may occur during examinations or quite a long period beforehand (depending upon the stakes of the test), potentially impacting upon learning, revision and test preparation

(Cassady, 2004; Smith, 2018). Although correlated, emotionality and worry have been shown to have distinct, independent associations with test performance – though the effect of the former is typically weaker (on average) than the latter (Hembree, 1988).

Why might test anxiety be linked to performance in high-stakes exams? Many popular explanations stem from attentional theories, where the arousal induced by examinations impairs cognitive performance (e.g. Ng & Lee, 2016). One such example is 'explicit monitoring theory', a phenomenon often applied to sports stars to explain why they may 'choke under the pressure' (Yu, 2015). In essence, the pressure of exams may lead young people to become too self-focused, putting more and more pressure upon themselves to the point that it becomes detrimental to their performance (Wine, 1971). Such feelings may not only occur during the examination itself, but also leading up to it (e.g. when revising). An alternative explanation comes from 'distraction' and 'attitudinal control' theories. These postulate that test anxiety leads young people to lose focus, and that their attention gets split between preparing for/answering the examination questions and unhelpful negative thoughts about failing the exam and the potential negative consequences. The latter takes up valuable working memory (Cassady, 2004; Ikeda et al., 1996), meaning young people will be operating below maximum capacity, with a negative effect upon their performance (Dutke & Stöber, 2001).

It is important to realise, however, that the effects of test anxiety upon academic performance have been disputed for at least two reasons. First, although it is often assumed that anxiety is a negative experience, the direction of its association with academic achievement is (at least theoretically) not entirely clear (Kader, 2016). Although empirical evidence over several decades has found there to be a negative correlation (Von der Embse et al., 2018) this is likely to depend on how anxious young people feel and how they react to stress. As noted by Howard (2020), whereas some individuals may be motivated by feelings of anxiety (e.g. thoughts like 'I best study hard for the upcoming exams or I might fail') for others it may interfere with cognitive processes meaning academic performance is reduced. Hence the association between test anxiety and exam grades may depend upon the amount of anxiety experienced. Arguably, very low levels of test anxiety in the build-up to high-stakes exams could be as bad (or even worse) for exam performance than high levels of anxiety, if it leads to a lack of sufficient preparation. Hence it may be that moderate levels of test anxiety are ideal, enough to stimulate motivation and spur young people into action, but without being so extreme as to interfere with cognitive processes and thus distracting young people from the task at hand. This is, in other words, an application of Yerkes-Dodson's law (Teigen, 1994) where very high and very low 'arousal' responses in the face of examinations are linked to lower levels of academic performance, compared to those in the middle of the distribution with 'average' anxiety levels.

A second challenge to there being a negative relationship between test anxiety and academic achievement comes from the work of Tobias (1985, 1990) and Sommer and Arendasy (2014). Such authors argue that there are two possible explanations for why such a negative association has been observed. The first – which they label as the 'interference model' – implies test anxiety has a causal effect upon achievement, interfering with students' ability to recall prior learning in an examination environment (or to retain such information effectively when revising). This follows the same logic as the paragraphs outlined above. In contrast, the deficit hypothesis '*claims that test anxiety and*

test performance are correlated because less competent test-takers experience higher levels of state test anxiety in the assessment process' (Sommer & Arendasy, 2014). Under the deficit model, a negative correlation between test anxiety and achievement is hence observed largely due to confounding/selection – less able pupils are more anxious about how they will perform in upcoming tests – and does not represent a causal relationship. Although some studies have presented empirical evidence consistent with the deficit model (e.g. Sommer & Arendasy, 2014) it has been widely recognised that teasing apart such 'selection-upon-ability' effects in the test anxiety-achievement relationship is difficult (Howard, 2020; Von der Embse et al., 2018).

Research questions

The theoretical background and literature presented above leads to the following research questions. First, the link between the 'worry' component of test anxiety and pupils' achievement in their GCSEs is explored. In doing so, the paper pays particular attention to the concerns of the deficit hypothesis, and that any apparent negative relationship could simply be due to academic selection (i.e. less able pupils experiencing higher levels of worry about the test). Specifically, I investigate the magnitude of the relationship between test anxiety and achievement on the high-stakes GCSE exams, under various different assumptions about how such selection effects are best controlled. This, in turn, gives credible upper and lower bounds on the link between test anxiety and GCSE performance. Thus, in summary:

- Research question 1. Is there a negative association between test anxiety and GCSE achievement, and to what extent is this due to 'selection effects' that are consistent with the deficit model?

Next, I turn to the issue of non-linearities. As noted above, it has been widely recognised that some degree of test anxiety may not be problematic, and may actually be positively related to achievement (to the extent that it helps motivate young people to study and thoroughly prepare). Yet, once anxiety passes a certain point, sizeable negative effects might be observed, due to young people not being able to perform at optimum capacity. There are hence clear theoretical reasons why one might anticipate the relationship between test anxiety and GCSE achievement to be non-linear, consistent with Yerkes-Dodson's law. However, few existing studies have explicitly considered this issue. The second research question addresses this point by asking:

- Research question 2. Is the association between test anxiety and GCSE performance non-linear? Is there evidence that particularly low or particularly high levels of test anxiety have an especially strong negative link to achievement?

Finally, I also consider differences between sub-groups. This includes variation by socio-economic status and prior achievement. Previous research has explored such heterogeneous effects, though with mixed results (Putwain, 2008). Results from this paper will hence help further strengthen the evidence base as to whether test anxiety is more strongly associated with the academic achievement of some groups compared to others:

- Research question 3. Does the relationship between test anxiety and GCSE achievement vary by prior achievement and socio-economic background?

3. Data and methods

Sample design

PISA is an international study of 15-year-olds' achievement in reading, mathematics and science. This paper uses data from the 2015 cycle. To ensure national representivity, PISA uses a two-stage sample design. Schools are first sampled with probability proportional to size, with 42 pupils then randomly selected to take part within each school (OECD, 2016). Almost every pupil that participates in PISA in England is in Year 11 (the school grade in which GCSE exams are taken). In total, 5,194 pupils from across 206 schools in England participated in PISA 2015. This equates to a school-level response rate of 92% and a pupil response rate of 88%. Final pupil response weights are applied throughout the analysis.

The PISA 2015 sample for England has been linked to the National Pupil Database (NPD). In total, a link between PISA and the NPD was made for 4,914 pupils (95% of the full sample).[1] An important feature of this survey-administrative linked database is that it includes information on how young people performed on two assessments (PISA tests and GCSE examinations) taken just six months apart.

Test anxiety

The PISA 2015 student questionnaire included the following questions reported using a four-point scale (strongly agree to strongly disagree):

- I often worry that it will be difficult for me taking a test.
- I worry that I will get poor grades at school.
- Even if I am well prepared for a test I feel very anxious.
- I get very tense when I study for a test.
- I get nervous when I don't know how to solve a task at school.

These questions capture the 'worry' aspect of test anxiety (e.g. fear of failure) rather than 'emotionality' (i.e. no question was asked about physiological symptoms – such as feeling nauseous – during tests). See, Borgonovi and Pal (2016) for further details. Note that the questions ask about young people's feelings about examinations and schoolwork in general, and not specifically about their experiences during the PISA test. Given the timing of PISA in England – six months before GCSEs – young people will likely be answering these questions with these upcoming high-stakes examinations in mind. As this question was asked at a single point in time, we are unable to explore the stability in test anxiety over the course of Year 11, and whether measurement at a later point (closer to when GCSEs are taken) might lead to different results. Previous research has suggested that pupil's 'state test anxiety' tends to slightly increase in the course of a semester as the final examination draws near, but with an assumption that 'trait test anxiety' (of which worry is a key component) is '*a stable personal disposition*' (Lotz & Sparfeldt, 2017).

The survey organisers have used students' responses to these questions to create a test anxiety scale (OECD, 2016). This scale has been reported to have good levels of reliability (Cronbach's alpha = 0.85 for the UK – OECD, 2016) and has been previously used in research exploring the correlates of test anxiety amongst 15-year-olds (e.g. Govorova, Benítez and Muñiz 2020). This is the covariate of interest in this paper, with the distribution for the English PISA sample presented in Figure 1 panel (a).[2]

Table 1 provides further details on the distribution of responses to each test anxiety question. This illustrates how the questions underpinning the test anxiety scale has a high degree of variability, with pupils providing responses across all four points of the 'strongly agree' to 'strongly disagree' response scale.

Outcome measure

Our primary outcome is young people's GCSE examination grades. These are the first set of high-stakes examinations that young people in England sit, reflecting the knowledge and skills they have accumulated throughout secondary school. How young people perform in these exams is important for their future educational and labour market opportunities. They are, for instance, used by sixth-form colleges and universities as part of their entrance criteria, while achieving certain GCSE grades is a requirement of many employers. These exams are taken by almost all pupils in May/June of Year 11. In total, students usually take GCSEs in around eight or more subjects, often involving around 20 examinations or more (totalling 30 hours of examination time or more). These examinations are widely seen as stressful for young people (McCaldin et al., 2019) given (a) the potential consequences of failure and (b) the intensity of the examination schedule. It is, consequently, an ideal environment to conduct test anxiety research.

Young people's capped total GCSE points score is the outcome variable of interest. This captures the grades achieved across eight subjects (with mathematics double weighted) and is a widely used summary measure of overall GCSE performance (Department for Education, 2015). The distribution of this measure for the analytic sample can be found in Figure 1 panel (b). Note that this measure has been standardised to mean zero and standard deviation one, meaning all results can be interpreted in terms of effect sizes. Grades achieved in GCSE mathematics is used as an alternative outcome measure to test the sensitivity of results, with results reported as grade differences (see Appendix B for further details).

Prior ability/achievement

Two measures of prior achievement are used in this paper – Key Stage 2 scores and PISA scores.

Key Stage 2 scores refer to the tests children sit at the end of primary school, when they are 10/11-years-old. They primarily cover skills in English and mathematics, though with a teacher-assessed grade also available in science. For the purposes of this paper, Key Stage 2 scores have both strengths and weaknesses. On the one hand, they are high-quality standardised assessments that are routinely used in academic research in England. They are not directly high stakes for children, in that poor performance does not have major material consequences for them as individuals. Moreover, the fact that they are

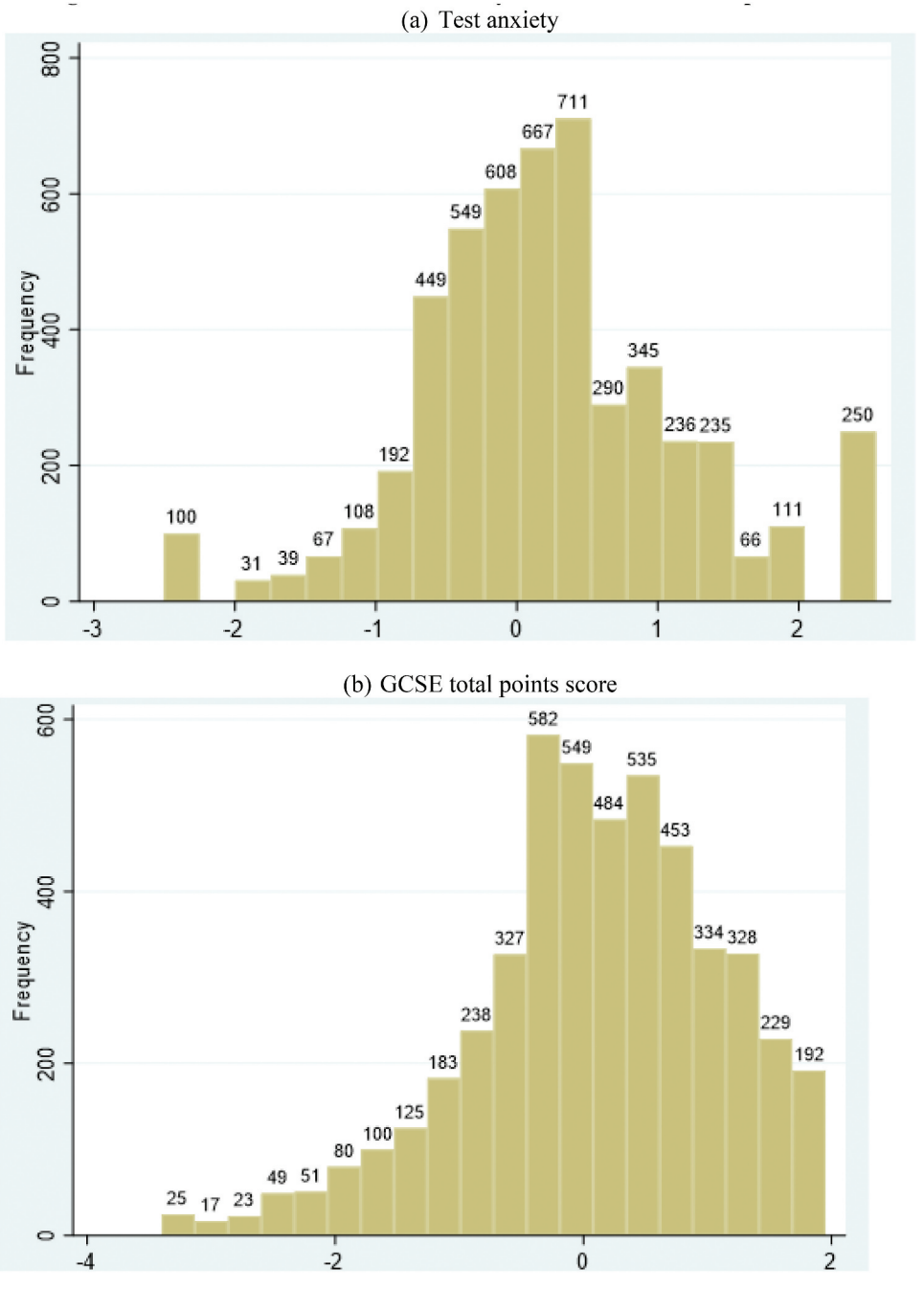


Figure 1. The distribution of the test anxiety scale and GCSE total points scores. (a) Test anxiety. (b) GCSE total points score. Both measures have been standardised to mean zero and standard deviation one. The figures running along the horizontal axis hence refer to the standardised values.

Table 1. The distribution of responses to each of the test anxiety items.

	Strongly disagree	Disagree	Agree	Strongly agree
I often worry that it will be difficult for me taking a test.	8% (404)	30% (1,535)	47% (2,339)	16% (774)
I worry that I will get poor grades at school.	6% (343)	27% (1,370)	44% (2,201)	23% (1,122)
Even if I am well prepared for a test I feel very anxious.	7% (351)	22% (1,104)	47% (2,336)	25% (1,235)
I get very tense when I study for a test.	9% (491)	38% (1,932)	37% (1,844)	15% (745)
I get nervous when I don't know how to solve a task at school.	10% (540)	35% (1,789)	38% (1,896)	17% (811)

Percentages refer to weighted figures. Figures in brackets refer to the unweighted number of observations.

taken five years before GCSEs – when children are still quite young – may mean that (arguably) scores on these tests are unlikely to be significantly affected by test anxiety (at least not to the same extent as GCSEs).

On the other hand, Key Stage 2 scores also have clear limitations. As some secondary schools use Key Stage 2 scores to decide pupils set/stream placement (and potentially the tier of the GCSE paper they sit) they may have some consequences for some pupils. They are also high-stakes for schools who get publicly ranked in 'league tables' based upon pupils' results. Hence many children are reportedly under pressure from their schools and teachers to perform well. Pupils may also not recognise the fact that these tests are high-stakes for their school rather than for themselves. Relatedly, there have been some reports suggesting that primary school children are anxious and 'feel stressed' about the Key Stage 2 tests (Hutchings, 2015; Reay & Wiliam, 1999), suggesting that they may to some extent also be impacted by test anxiety. It might also be that younger, less mature pupils could be less emotionally developed and prepared to take tests than older pupils. Finally, they are a measure of prior achievement recorded a long time (five years) before young people sit their GCSEs. Although the Key Stage 2-GCSE correlation is reasonably strong (Pearson correlation of 0.75 for the analytic sample in reference to mathematics), Key Stage 2 scores alone will not capture and control for the learning gains young people will have made during secondary school.

The second measure of prior achievement used is PISA scores. These capture Year 11 pupils' skills in reading, mathematics and science six months before they take their GCSEs in England. As a control for prior achievement in the build-up to GCSE exams, they have several attractions. First, they correlate reasonably well with GCSE grades (e.g. Pearson $r = 0.68$ for the correlation between PISA and GCSE mathematics). Second, they are very low stakes. There are no consequences for students or schools depending upon how they perform; in fact, pupils do not find out their results. This is clearly in direct contrast to the very high-stakes nature of GCSEs, meaning PISA scores should (arguably) be a lot less affected by test anxiety than GCSE grades.

PISA scores do, however, also have some limitations. The low-stakes nature of the PISA test could lead to sub-optimal levels of test effort, meaning that they do not fully capture young people's true level of skill (Gneezy et al., 2019). There are also important subtle differences in the PISA and GCSE test constructs, with the former capturing teenagers' 'functional abilities' (i.e. how well they can apply skills to solve 'real world' problems) while the latter is designed to measure young people's mastery of England's national curricula (Carvalho, 2010).

The methodology sub-section that follows below discusses the impact these features of the PISA and Key Stage 2 tests have upon the interpretation of the results.

Measurement of socio-economic status

Within our analysis, we explore whether there is differential impact of test anxiety by socio-economic group. Socio-economic status is measured using the PISA Economic and Social Cultural Status (ESCS) scale. As part of the background questionnaire, pupils are asked about their mothers' and fathers' education, occupation and household possessions. The survey organisers create a continuous scale using this information via principal components analysis.[3]

Methodology

Year 11 students are divided into deciles of the test anxiety scale, with the middle (fifth) decile as the reference group. This is a straightforward way of capturing potential non-linearities in the test anxiety-GCSE achievement relationship (thus addressing research question 2) using standard and widely understood statistical techniques.

Formally, estimates will be presented from a series of OLS regression models:

$$GCSE_{ij} = \alpha + \beta.Anx_i + \delta.D_i + \gamma.A_i + \mu_j + \varepsilon_{ij} \quad (1)$$

Where:

Anx_i = A vector of dummy variables capturing declines of the test anxiety scale.

D_i = A vector of demographic characteristics (socio-economic status, parental education).

A_i = Prior achievement measured by PISA/Key Stage 2 scores.

μ_j = School fixed-effects. (Appendix F tests the robustness of results to removing these from the model).

ε_{ij} = Error term

i = Student i

j = School j.

Six specifications of this model are presented in the main text. These are mainly based around different assumptions about how prior achievement is controlled.

The baseline model (M0) is the 'empty' model which does not include any controls. The first model specification (M1) adds controls for socio-economic background, given that this factor might confound the relationship between test anxiety and GCSE grades. For instance, previous research has found that students from disadvantaged backgrounds have higher levels of test anxiety (Putwain, 2007, 2008), while socio-economic status is also a strong predictor of GCSE grades. Controlling for socio-economic status is hence important to ensure it does not drive any apparent link between test anxiety and GCSE performance. In model M2, school fixed-effects are added (though prior achievement is still not controlled). Results from these first specifications hence ignore the 'deficit' hypothesis, with an assumption that there is effectively no selection-by-ability (i.e. that less able pupils are no more or less likely to suffer test anxiety than more able pupils). This is, of course, quite a strong assumption to make, with prior research suggesting it unlikely

to hold (Sommer & Arendasy, 2014). Estimates from models M0, M1 and M2 will, however, provide a credible upper-bound on the strength of the association between test anxiety and GCSE grades.

This assumption is relaxed in the following two model specifications, where *either* PISA scores (M3) or Key Stage 2 scores (M4) are controlled. These models recognise (and account for) the fact that pupils with lower levels of prior achievement may be more anxious about upcoming GCSE exams. The key assumption now being made is that scores on these tests (either Key Stage 2 or PISA) are largely unaffected by test anxiety. Previous research has indeed suggested that 'student test anxiety is higher on high-stakes exams when compared with typical classroom tests' (Von der Embse et al., 2018) and that 'students are overwhelmed with stress, anxiety, and worry due to testing in high-stakes contexts' (Silaj et al., 2021). However, as previously noted, the assumption that Key Stage 2 or PISA tests are unaffected by test anxiety is likely only approximately to hold true. To the extent that Key Stage 2 or PISA scores have been impacted by test anxiety, results from M3 and M4 may lead to underestimation of the strength of the association between test anxiety and GCSE performance. Nevertheless, model M3 (controlling for PISA scores only) is the preferred specification (and provides the headline results), due to it providing a strong control for prior achievement taken just six months before GCSE which is unlikely (due to its very low-stakes nature) to have been substantially affected by test anxiety.

A similar argument holds for model M5, where *both* Key Stage 2 scores and PISA scores are controlled. By controlling for both measures, model M5 includes particularly rich measures of young people's prior achievement. On the other hand, a stronger assumption must be made – that *neither* Key Stage 2 nor PISA scores have been impacted by test anxiety – if one is to interpret the estimates as capturing the effect of test anxiety upon GCSE performance. This may only approximately hold true. Yet estimates from this final model specification are likely to still be useful, in that they arguably provide a credible lower-bound for the association between test anxiety and GCSE grades.

For all model specifications, multiple imputation has been used to account for missing covariate data. Standard errors have been clustered at the school-level to account for the clustering of pupils within schools. The first plausible values in mathematics, reading and science have been included in the models that control for PISA scores, though with the substantive results largely unchanged if all ten plausible values are used instead (see Jerrim et al., 2017 for further details).

After estimating the model using the full sample, separate results are produced for the following sub-groups:

- Low socio-economic status (bottom ESCS quartile)
- High socio-economic status (top ESCS quartile)
- Low achievement (bottom PISA science quartile)
- High achievement (top PISA science quartile)

Finally, a series of robustness tests will be conducted to explore the sensitivity of results, reported in detail in the online appendices. First, non-parametric regression is used (rather than Ordinary Least Squares) to estimate the models. This provides an alternative way to explore potential non-linearities in the relationship between test anxiety and GCSE performance. Second, rather than using total GCSE points score as the outcome variable

of interest, GCSE mathematics grades are used instead. Third, in England, there is often interest in children who sit on key grade boundaries (e.g. a C grade in English and mathematics) as falling one side of the grade boundary versus another has significant consequences for future lifetime outcomes. These young people may hence also be particularly anxious about how they perform in their GCSEs. Consequently, in Appendix D, separate estimates are presented for this sub-group (defined as those who achieved either a C or D grade in mathematics). Finally, an alternative set of estimates are presented where the controls included are either expanded or reduced. This helps to illustrate the sensitivity of results to the precise model specification.

4. Results

Primary analysis

The headline results can be found in Figure 2. This can be cross-referenced with Table 2 which presents the full set of parameter estimates. The horizontal axis plots deciles of the test-anxiety scale, running from low (decile 1–10) to high (decile 91–100) levels of anxiety. The vertical axis illustrates the difference in GCSE outcomes between young people within each test-anxiety decile, with the fifth (41–50) decile as the reference group. Higher (lower) values indicate better (worse) outcomes, with estimates presented as effect sizes in panel (a) (total GCSE point scores) and grade differences in panel (b) (GCSE mathematics). Results from three specifications are presented: (a) M2, where just demographics and school fixed-effects are controlled (upper-bound on effect); (b) M3, where PISA scores are added to the model (central estimate) and (c) M5 where both PISA and Key Stage 2 scores are controlled (likely lower-bound on the effect).

The first notable feature of Figure 2 is that there is very little difference in GCSE outcomes between young people with 'typical' levels of test anxiety and those who are at the top-end of the test anxiety scale. This holds true across all three of the model specifications presented in Figure 2, with the lines (representing the estimated effect) being essentially flat between the 41—50 and 91—100 decile. In terms of effect sizes, all estimates between the 41—50 and 91—100 test anxiety decile are below 0.1 (and typically sit very close to zero). There is hence little evidence that suffering from high levels of test anxiety is detrimental to GCSE performance; young people who are very anxious about testing/examinations achieve the same GCSE grades as their peers with 'typical' levels of test anxiety.

There is somewhat more nuance to this result when comparing differences in GCSE outcomes between young people with very low levels of test anxiety (e.g. 1–10 and 11–20 decile) and those with 'typical' levels (e.g. 41–50 decile). In the model including only demographic background controls, there is a moderate, positive effect size (≈ 0.2) when comparing children with low levels of test anxiety to those with moderate/high levels. This, however, falls considerably in the model specifications where prior achievement (and, most notably, PISA scores) have been controlled. Indeed, after accounting for PISA scores, effect sizes fall below 0.1 standard deviations, with the lines plotted in Figure 2 now essentially flat across the test anxiety distribution. This is further emphasised by Table 2, with all estimates from model specification 3 (PISA scores controlled) and specification 5 (PISA and

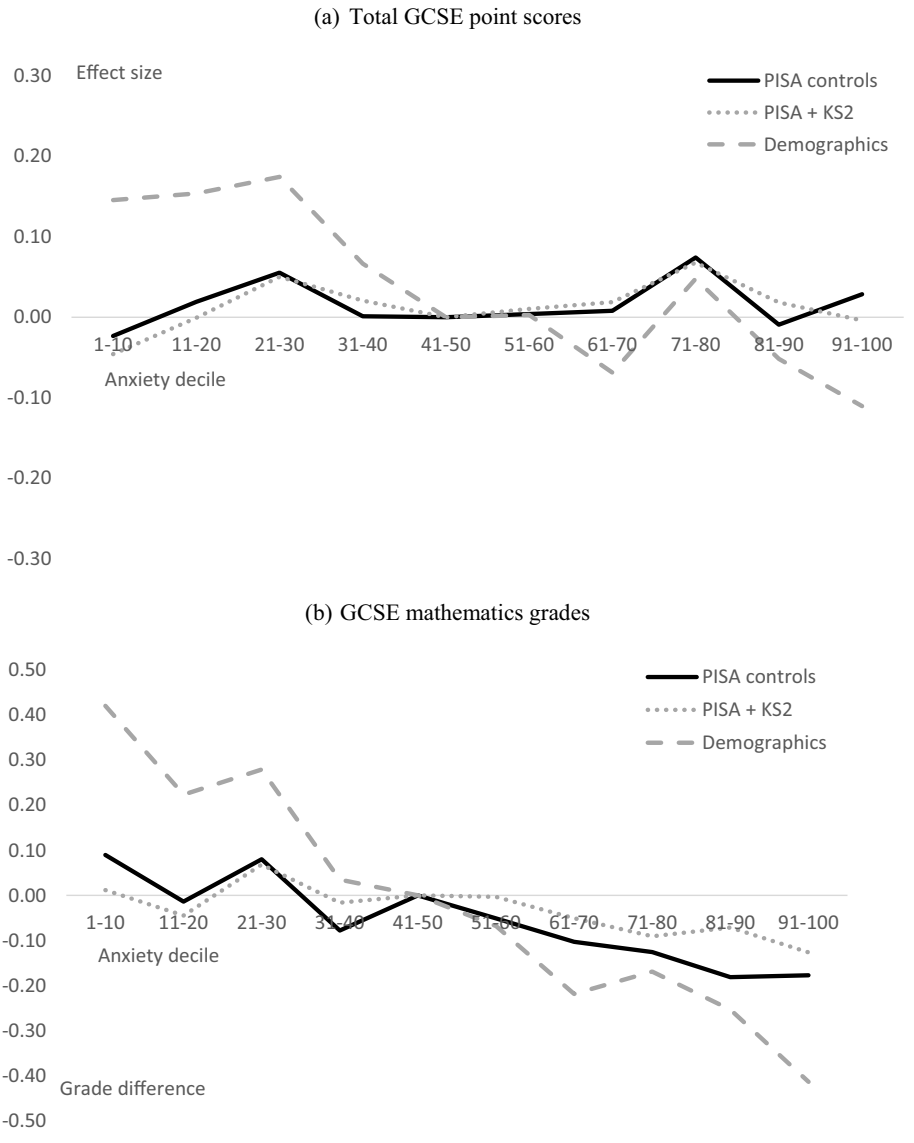


Figure 2. The link between test-anxiety decile and GCSE grades. (a) Total GCSE point scores. (b) GCSE mathematics grades. Figures in panel (a) refer to effect sizes, while those in panel (b) refer to differences in grades. Graph can be cross-referenced with Table 2 (with additional details on model specifications provided in the table notes). Positive (negative) values indicates higher (lower) GCSE grades. 'Demographics' refer to estimates from Model M2 (demographics and school fixed-effects controlled). 'PISA controls' refer to estimates from Model 3 (demographics, school fixed-effects and PISA scores controlled). 'PISA + KS2' refer to estimates from Model M5 (demographics, school fixed-effects, PISA scores and Key Stage 2 scores controlled).

Key Stage 2 scores) small (effect sizes mostly below 0.1) and typically statistically insignificant at conventional thresholds. Thus, with respect to the first research question, there is little evidence of a clear link between test anxiety and GCSE performance.

Table 2. The link between test-anxiety decile and GCSE grades (parameter estimates).

(a) GCSE total point scores (effect sizes)

	Q1		Q2		Q3		Q4		Q6		Q7		Q8		Q9		Q10	
Model	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE
0	0.14*	0.07	0.18*	0.06	0.19*	0.07	0.06	0.07	−0.01	0.07	−0.16*	0.08	0.00	0.07	−0.09	0.08	−0.18*	0.07
1	0.12	0.06	0.12	0.06	0.17*	0.06	0.05	0.07	−0.02	0.07	−0.12	0.07	−0.01	0.06	−0.06	0.07	−0.14*	0.06
2	0.15*	0.06	0.15*	0.06	0.17*	0.06	0.07	0.06	0.00	0.06	−0.07	0.06	0.05	0.06	−0.05	0.07	−0.11	0.06
3	−0.02	0.05	0.02	0.04	0.06	0.04	0.00	0.05	0.00	0.04	0.01	0.05	0.07	0.05	−0.01	0.05	0.03	0.05
4	0.00	0.04	0.04	0.04	0.09	0.05	0.06	0.04	0.02	0.04	0.00	0.04	0.06	0.04	0.02	0.05	−0.07	0.05
5	−0.05	0.04	0.00	0.04	0.05	0.04	0.02	0.04	0.01	0.04	0.02	0.04	0.07	0.04	0.02	0.04	0.00	0.04
N	**547**		**512**		**506**		**507**		**492**		**504**		**512**		**496**		**487**	

b) GCSE mathematics (grade differences)

Model	Q1		Q2		Q3		Q4		Q6		Q7		Q8		Q9		Q10	
	Grade diff	SE	Grade diff	SE	Grade diff	SE	Grade diff	SE	Grade diff	SE	Grade diff	SE	Grade diff	SE	Grade diff	SE	Grade diff	SE
0	0.38*	0.11	0.24*	0.11	0.28*	0.11	0.03	0.12	−0.08	0.13	−0.40*	0.14	−0.28*	0.12	−0.30*	0.14	−0.55*	0.10
1	0.37*	0.10	0.16	0.10	0.26*	0.11	0.02	0.11	−0.07	0.11	−0.32*	0.12	−0.27*	0.10	−0.25*	0.12	−0.50*	0.09
2	0.42*	0.10	0.22*	0.10	0.28*	0.10	0.04	0.10	−0.07	0.11	−0.22*	0.11	−0.17	0.10	−0.25*	0.12	−0.41*	0.09
3	0.09	0.08	−0.01	0.08	0.08	0.08	−0.08	0.08	−0.05	0.08	−0.10	0.09	−0.13	0.08	−0.18	0.08	−0.18*	0.07
4	0.10	0.07	0.02	0.07	0.13	0.08	0.05	0.07	0.00	0.08	−0.07	0.08	−0.11	0.08	−0.06	0.08	−0.23*	0.07
5	0.01	0.07	−0.04	0.07	0.07	0.07	−0.02	0.06	0.00	0.07	−0.05	0.07	−0.09	0.07	−0.07	0.07	−0.13	0.06
N	**476**		**461**		**449**		**448**		**449**		**457**		**461**		**452**		**441**	

Decile 5 (average levels of test anxiety) forms the reference group. Decile 1 = least anxious and decile 10 = most anxious. Figures for GCSE total points scores in panel (a) refer to effect sizes, while figures for GCSE mathematics grades in panel (b) refer to differences in grades (e.g. a figure of 0.3 in panel (b) refers to around one-third of the GCSE mathematics grade). * indicates statistically significant difference from the reference group at the five per cent level. Model 0 does not include any controls. Model 1 controls for demographics. School fixed-effects are added in Model 2. PISA scores are added as controls in Model 3. Key Stage 2 scores (but not PISA scores) are added in Model 4. Model 5 includes both PISA scores and Key Stage 2 scores (along with demographics and school fixed-effects) as controls.

The results presented in Figure 2 also provide the evidence needed to address research question 2. Here, it was hypothesised that the relationship between test anxiety and GCSE grades may be non-linear, with strong *negative* effects on performance for those at both the top and the bottom ends of the test anxiety scale. Were this true, the lines plotted in Figure 2 should form an 'n' shape; young people in the 1—10, 11—20, 81—90 and 91—100 test anxiety deciles would achieve lower GCSE grades than those in the middle (e.g. 41–50, 51–60 percentiles) of the test anxiety distribution. Clearly, this is not the case. There is no evidence of strong non-linearities in any of the model specifications. Indeed, in the models that do not control for prior achievement, the effect at the lower end of the test anxiety distribution is actually in the opposite direction (i.e. those with very low levels of test anxiety actually achieve higher – not lower – grades than those in the middle of the test anxiety distribution). There is hence sufficient evidence to reject Yerkes-Dodson's law holding in this particular setting.

Sub-group results

Appendix G presents analogous graphs for our sub-groups of interest: low socio-economic status (panel a), high socio-economic status (panel b), low-achieving pupils (panel c) and high-achieving pupils (panel d). On the whole, there is no clear evidence of differences between any of these sub-groups; the same general pattern can be observed in each graph. In particular, for all groups, there is no substantive difference in GCSE outcomes for highly test-anxious pupils (in comparison to 'typical' young people with average levels of test anxiety). Thus, in general, there is no clear pattern that test anxiety is linked to GCSE performance for any of these key subgroups.

Item-level analysis

Appendix H provides the item-level analysis, focusing upon the association between how responses to each of the five test-anxiety questions are associated with GCSE grades. (See Appendix A for the percentage of 15-year-olds who provided each response to each question.) Focusing upon the results using the preferred (M3) model specification (illustrated using the black line) for all the questions bar one there is clearly no link with GCSE outcomes. This should not be surprising, given the results reported in the previous sub-sections. The one exception is with respect to the statement: 'I worry that I will get poor grades at school'. For this question there is a relatively shallow, linear negative association. In particular, those who strongly disagree with this statement achieve around 0.1 standard deviations higher total GCSE point scores those their peers who strongly agree. Yet, with this single exception, the item-level analysis presented in Appendix H supports the conclusion that test anxiety is not related to the grades young people achieve in their GCSEs.

Sensitivity analyses

A series of robustness tests are provided in the online supplementary material. An overview of the conclusions drawn from these additional analyses are provided here.

Table 2b and Appendix B presents alternative estimates using mathematics GCSE grades – rather than total GCSE point scores – as the outcome. (Note that results for GCSE maths grades are reported in a different metric – grade differences along with nine-point A*-U grade scale – rather than effect sizes[4].) Results and substantive conclusions reached are similar to those when using total GCSE point scores. Rather than dividing the test anxiety scale into deciles, Appendix C presents results using quintiles instead. Again, this does not lead to any material change to the results. In Appendix D the sample is restricted to those young people on the key C/D grade boundary in mathematics,[5] with linear probability models estimated for this key sub-sample. There is again no evidence that test anxiety is strongly associated with the GCSE achievement of this group. Appendix E uses non-parametric regression, rather than OLS, to estimate the models.[6] Consistent with the findings reported above, there is no evidence of strong non-linearities in the test-anxiety GCSE achievement relationship. Finally, Appendix F alters the set of the controls in the analysis models, either by removing the school fixed-effects, or adding additional variables controlling for other aspects of young people's socio-emotional state (e.g. their motivation, future aspirations etc). There is no evidence that the key findings reported above are sensitive to the choice of controls included.

5. Conclusions

The mental health and wellbeing of young people has become an issue of much public policy attention (Buck & Woods, 2019). Although this is a complex area, the stress and anxiety induced by high-stakes assessment is thought by many to be a significant problem amongst teenagers in England (Putwain & Daly, 2014). Such anxiety about testing is not only important for young people's mental health and wellbeing, but also potentially for their educational achievement. Indeed, a wide-ranging literature has noted how there is a negative correlation between test anxiety and test performance, with those individuals who are less worried tending to achieve higher school grades than their more anxious peers (Von der Embse et al., 2018). There has, however, been some debate as to whether this reflects a causal relationship (Sommer & Arendasy, 2014). In particular, proponents of the 'deficit hypothesis' argue that this negative relationship is being driven (at least in part) by less academically able and prepared young people being more anxious about important, upcoming tests (Tobias, 1985, 1990). The strength of the association between test anxiety and examination performance may thus be overestimated unless this issue is considered (Sommer & Arendasy, 2014). Moreover, surprisingly little work has considered the extent to which the link between test anxiety and examination performance is non-linear. Specifically, is there a particular point when test anxiety becomes too much and examination performance starts to decline? Similarly, is there any evidence that test anxiety can be too low, with a casual attitude and complacency about important upcoming examinations meaning some teenagers end up achieving worse results?

This paper has provided new evidence on such issues for Year 11 pupils in England, as they approach the high-stakes GCSE exams. Using PISA 2015 data linked to national administrative records, the analysis has explored the strength of the association between test anxiety and the GCSE grades young people achieve, and how this changes depending

upon whether (and how) prior achievement is controlled. Non-linear effects have been explored, thus illustrating whether GCSE outcomes differ between those teenagers who report very high (or very low) levels of test anxiety, and those whose anxiety levels are in a more 'normal' range.

On the whole, the paper has presented evidence of largely null effects. There is little difference in GCSE grade outcomes between young people with high levels of test anxiety compared to their peers in the middle of the test anxiety distribution. This holds true across different model specifications, using different outcome measures and survives a series of robustness tests. Similarly, no evidence is found of significant non-linearities in the test anxiety-GCSE performance relationship, rejecting the notion that Yerkes-Dodson's law applies in this context. There is also no suggestion of meaningful differences across key sub-groups (based upon prior achievement and socio-economic status), with no clear evidence emerging of heterogeneous effects.

These findings are in contrast to the general thrust of the literature to date where, by-and-large, moderately-sized negative effects have been found. It is noteworthy, however, that the negative correlations reported in widely cited meta-analyses are either unconditional estimates or only conditional upon a relatively small set of potential confounding factors (Von der Embse et al., 2018). Indeed, a handful of other studies in this literature have also found there to be only very small effects. This is particularly true once there has been some attempt to control for potential selection bias induced by academically weaker pupils developing higher levels of anxiety (Sommer & Arendasy, 2014). Moreover, the results presented in this paper are consistent with the deficit hypothesis – in that the negative correlation between test anxiety and GCSE examination performance seems largely due to confounding/selection.

These findings should of course be interpreted in light of the limitations on this research. First, the measure of test anxiety used is geared more towards the 'worry' aspect of test anxiety rather than 'emotionality'. It also comprises five items, compared to other test anxiety scales which are longer (e.g. 20 items for the Test Anxiety Inventory used in Putwain, 2007). This is a natural consequence of conducting a secondary analysis of the general-purpose PISA dataset, rather than bespoke data collection that focuses solely upon the issue of test anxiety. Nevertheless, differences in measurement could explain some of the contradictory research findings, with replication of this study encouraged using other datasets and measurement tools. Second, relatedly, test anxiety has been measured at just one single point in Year 11. Ideally, measurement of test anxiety across multiple points during the academic year would facilitate a richer analysis. This would develop our understanding both in terms of how test anxiety changes as the high-stakes GCSE examinations draw nearer (i.e. the extent that 'worry' and 'emotionality' are stable traits) and whether the timing of its measurement matters for its association with GCSE grades. Although this is a demanding data requirement – particularly for a nationally representative sample – it should be considered an important next step in this line of research. Third, as part of England's GCSE exams, special measures are in place for examination candidates with extenuating circumstances, including mental health problems which encompasses those with anxiety issues (Ofqual, 2019). Unfortunately, information about such special measures is not available within the data held. Yet this could provide important context to the results, potentially suggesting that the mitigating strategies currently in place (such as extra time provided to candidates) are to some

extent working. Finally, the empirical analysis has been conducted within a specific context (England) for one particular year group (Year 11). Given the high-stakes nature of GCSEs – and the sheer number of examinations 15/16-year-olds take over a six-week period – this may be quite an atypical setting. The extent to which the findings generalise to other countries and/or age groups remains somewhat of an unknown.

Despite these limitations, results from this study have some important implications. One interpretation is that, whatever mitigating steps are currently in place to help test-anxious pupils through GCSEs, they seem to be 'working' (at least in terms of limiting any detrimental effect upon the grades young people achieve). Alternatively, our findings may suggest that pupils with very high levels of test anxiety perhaps ration their efforts, focusing upon not allowing their anxiety to disturb their studies, potentially to the detriment of their mental health. Indeed, relatively little is known about how young people's wellbeing and mental health are causally affected by GCSE examinations, particularly in the form of large-scale quantitative research. Indeed, as suggested by an anonymous referee, exceptionally high levels of test anxiety could even lead to mental health issues that prevent students from taking GCSEs at all. Hence, while there seems little need to encourage further intervention from an educational achievement perspective, the same may not hold true with respect to teenagers' mental health. Future work in this area should therefore focus upon the extent that test anxiety is linked to the broader mental wellbeing of young people, and both the short- and long-term effects that this has on their lives.

Notes

1. Around 1.5% of the English PISA are Year 10 pupils and will not have had their GCSEs matched into the file. Independent school pupils were also less likely to have linked data on GCSE outcomes available. This is likely to due independent school pupils being disproportionately likely to take alternative qualifications (e.g. International Baccalaureate, IGCSEs). Thus, on most occasions, non-linkage of administrative records is unlikely to be strongly linked to mental health issues per se.
2. In the regression analysis, a small amount of random noise has been added to the scale in order to smooth the distribution. This has been done to ensure there are not 'ties' when we divide the sample into test anxiety deciles.
3. Principal components analysis (PCA) is a statistical technique that aims to reduce a large set of variables into a smaller set of variables, while minimising information loss. In this context, the aim of the PCA was to reduce the large set of socio-economic status measures collected down into a single continuous socio-economic status scale. For an intuitive tutorial of PCA, see, Shlens (2005).
4. Results can be converted from GCSE grade differences into an effect size by dividing the estimates reported in Appendix B by 1.9.
5. In 2017 a new grading system was introduced in England, with numeric grades replacing the old alphabetical grades The key grade boundary in mathematics for pupils is now grade 4. The pupils included in this analysis took their GCSEs in 2016, before this change to the grading system took place.
6. School fixed-effects are not included in these models to facilitate convergence.

Acknowledgments

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. This research was supported by the ESRC. Research grant ES/T003677/1.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Economic and Social Research Council [ES/T003677/1].

Notes on contributor

John Jerrim is a Professor of Education and Social Statistics at UCL. He has a PhD from the University of Southampton, obtained in 2010. John has published widely on the issue of international comparisons of educational achievement, mainly using the PISA, TIMSS and PIRLS datasets. John led the national PISA reports for England, Northern Ireland and Wales in 2015, and the TALIS report in 2018. In his most recent projects, John has investigated the mental health and wellbeing of teachers and the links between high-stakes testing and young people's mental health.

ORCID

John Jerrim http://orcid.org/0000-0001-5705-7954

References

Borgonovi, F., & Pal, J. (2016). A framework for the analysis of student well-being in the PISA 2015 study. *OECD education working papers 140*. OECD publishing: Paris.

Buck, R., & Woods, K. (2019). *Understanding test anxiety*. The Ofqual Blog. Retrieved May 18, 2020, from https://ofqual.blog.gov.uk/2019/03/01/understanding-test-anxiety/

Carvalho, F. (2010). *A comparative investigation of England's national assessments and the international surveys*. Ofqual research report. Ofqual/10/4710. Retrieved May 19, 2020, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/604973/0510_FatimaCarvalho_Comparative-investigation-on-national-assessments.pdf

Cassady, J., & Johnson, R. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094

Cassady, J. (2004). The impact of cognitive test anxiety on text comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology*, *18*(3), 311–325. https://doi.org/10.1002/acp.968

Department for Education (2015). *Examination point scores*. Retrieved May 19, 2020, from https://dera.ioe.ac.uk/25970/1/Key%20stage%204%20and%2016%20to%2018%20examination%20point%20scores%202015.pdf

Doctor, R., & Altman, F. (1969). Worry and emotionality as components of test anxiety: Replication and further data. *Psychological Reports*, *24*(2), 563–568. https://doi.org/10.2466/pr0.1969.24.2.563

Dutke, S., & Stöber, J. (2001). Test anxiety, working memory, and cognitive performance: Supportive effects of sequential demands. *Cognition and Emotion*, *15*(3), 381–389. https://doi.org/10.1080/02699930125922

Everson, H., Millsap, R., & Browne, J. (1989). Cognitive interference or skills deficit: An empirical test of two competing theories of test anxiety. *Anxiety Research*, *1*(4), 313–325. https://doi.org/10.1080/08917778908248728

Gneezy, U., List, J., Livingston, J., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights, American Economic Association*, *1*(3), 291–308. https://www.aeaweb.org/articles?id=10.1257/aeri.20180633

Govorova, E., Benítez, I., & Muñiz, J. (2020). How schools affect student well-being: A cross-cultural approach in 35 OECD countries. *Frontiers in Psychology*, *11*, 431. https://doi.org/10.3389/fpsyg.2020.00431

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*(1), 47–77. https://doi.org/10.3102/00346543058001047

Howard, E. (2020). *A review of the literature concerning anxiety for educational assessments*. Ofqual research report. Retrieved January 20, 2021, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/865832/A_review_of_the_literature_concerning_anxiety_for_educational_assessment.pdf

Hutchings, M. (2015). *Exam factories? The impact of accountability measures on children and young people*. Report for the National Union of Teachers. Retrieved April 21, 2020, from https://www.basw.co.uk/system/files/resources/basw_112157-4_0.pdf

Ikeda, M., Iwanaga, M., & Seiwa, H. (1996). Test anxiety and working memory system. *Perceptual and Motor Skills*, *82*(3), 1223–1231. https://doi.org/10.2466/pms.1996.82.3c.1223

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, *61*(Suppl. C), 51–58. https://doi.org/10.1016/j.econedurev.2017.09.007

Kader, A. (2016). Debilitating and facilitating test anxiety and student motivation and achievement in principles of microeconomics. *International Review of Economics Education*, *23*, 40–46. https://doi.org/10.1016/j.iree.2016.07.002

Keogh, E., Bond, F., French, C., Richards, A., & Davis, R. (2004). Test anxiety, susceptibility to distraction and examination performance. *Anxiety, Stress & Coping*, *17*(3), 241–252. https://doi.org/10.1080/10615300410001703472

Lotz, C., & 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. https://doi.org/10.1016/j.paid.2016.08.032

McCaldin, T., Brown, K., & Greenwood, J. (2019). *What is it like to experience exam stress? A student perspective*. Ofqual blog. Retrieved May 19, 2020, from https://ofqual.blog.gov.uk/2019/03/08/what-is-it-like-to-experience-exam-stress-a-student-perspective/

Minor, S., & Gold, S. (1985). Worry and emotionality components of test anxiety. *Journal of Personality Assessment*, *49*(1), 82–85. https://doi.org/10.1207/s15327752jpa4901_15

Musch, J., & Broder, A. (1999). Test anxiety versus academic skills: A comparison of two alternative models for predicting performance in a statistics exam. *British Journal of Educational Psychology*, *69*(1), 105–116. https://doi.org/10.1348/000709999157608

Ng, E., & Lee, K. (2016). Test anxiety and children's working memory task performance: Does trait or state anxiety matter more? *Journal of Experimental Psychopathology*, *7*(3), 374–390. https://doi.org/10.5127/jep.054115

OECD. 2016. PISA 2015 technical report. OECD publishing:

Ofqual (2019). *Access arrangements for GCSE, AS and A level 2018 to 2019 academic year*. Ofqual research report Ofqual/19/6563/1. Retrieved May 19, 2020, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/847495/Access_Arrangements_for_GCSE_AS_and_A_level_2018_to_2019_academic_year.pdf

Putwain, D. (2007). Test anxiety in UK schoolchildren: Prevalence and demographic patterns. *British Journal of Educational Psychology*, *77*(3), 579–593. https://doi.org/10.1348/000709906X161704

Putwain, D. (2008). Test anxiety and GCSE performance: The effect of gender and socio-economic background. *Educational Psychology in Practice*, *24*(4), 319–334. https://doi.org/10.1080/02667360802488765

Putwain, D. W. (2009). Assessment and examination stress in Key Stage 4. *British Educational Research Journal*, *35*, 391–411. https://doi.org/10.1080/01411920802044404

Putwain, D., & Daly, A. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies*, *40*(5), 554–570. https://doi.org/10.1080/03055698.2014.953914

Putwain, D., & von der Embse, N. (2018). Teachers use of fear appeals and timing reminders prior to high-stakes examinations: Pressure from above, below, and within. *Social Psychology of Education*, *21*(5), 1001–1019. https://doi.org/10.1007/s11218-018-9448-8

Reay, D., & Wiliam, D. (1999). 'I'll be a nothing': Structure, agency and the construction of identity through assessment. *British Educational Research Journal*, *25*(3), 343–354. https://doi.org/10.1080/0141192990250305

Reeve, C., & Bonaccio, S. (2008). Does test anxiety induce measurement bias on cognitive ability tests? *Intelligence*, *36*(6), 526–538. https://doi.org/10.1016/j.intell.2007.11.003

Sawka-Miller, K. (2011). Test anxiety. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and development*. Springer. 1478.

Shlens, J. (2005). A tutorial on principal component analysis. Retrieved January 17, 2022, from https://www.cs.cmu.edu/~elaw/papers/pca.pdf

Silaj, K. M., Schwartz, S. T., & Siegel, A. L. M. (2021). Test anxiety and metacognitive performance in the classroom. *Educational Psychology Review*, *33*(4), 1809–1834. https://doi.org/10.1007/s10648-021-09598-6

Smith, C. (2018). *Test anxiety can be debilitating. But schools can help students manage it*. The Guardian 18th April 2018. Retrieved May 18, 2020, from https://www.theguardian.com/teacher-network/2018/apr/18/test-anxiety-can-be-debilitating-but-schools-can-help-students-manage-it

Sommer, M., & Arendasy, M. (2014). Comparing different explanations of the effect of test anxiety on respondents' test scores. *Intelligence*, *42*(1), 115–127. https://doi.org/10.1016/j.intell.2013.11.003

Sommer, M., & Arendasy, M. (2015). Further evidence for the deficit account of the test anxiety–test performance relationship from a high-stakes admission testing setting. *Intelligence*, *53*, 72–89. https://doi.org/10.1016/j.intell.2015.08.007

Sung, Y. T., Chao, T.-Z., & Tseng, F.-L. (2016). Re-examining the relationship between test anxiety and learning achievement: An individual-differences perspective. *Contemporary Educational Psychology*, *46*, 241–252. https://doi.org/10.1016/j.cedpsych.2016.07.001

Teigen, K. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, *4*(4), 525–547. https://doi.org/10.1177/0959354394044004

Tobias, S. (1985). Test anxiety: Interference, defective skills and cognitive capacity. *Educational Psychologist*, *20*(3), 135–142. https://doi.org/10.1207/s15326985ep2003_3

Tobias, S. (1990). *Test anxiety: Cognitive interference or inadequate preparation*. Retrieved May 05, 2020, the American Educational Research Association, from https://www.researchgate.net/publication/235120899_Test_Anxiety_Cognitive_Interference_or_Inadequate_Preparation

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, *227*, 483–493. https://doi.org/10.1016/j.jad.2017.11.048

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, *76*(2), 92–104. https://doi.org/10.1037/h0031332

Yu, R. (2015). Choking under pressure: The neuropsychological mechanisms of incentive-induced performance decrements. *Frontiers in Behavioral Neuroscience*, *9*, 19. https://doi.org/10.3389/fnbeh.2015.00019