Test anxiety in Primary School Children: A 20 year Systematic Review and Meta-Analysis

Davina A.Robson[1]

[1]School of Psychology, University of Wollongong, Wollongong, Australia

Stuart J. Johnstone[1]

[1]School of Psychology, University of Wollongong, Wollongong, Australia

David W. Putwain[2]

[2]School of Education, Liverpool John Moores University, Liverpool, UK.

Steven Howard[1]

[1]School of Education, University of Wollongong, Wollongong, Australia

Author Note

Davina A. Robson https://orcid.org/0000-0002-9884-6746

Stuart J. Johnstone https://orcid.org/0000-0001-5380-9952

David W. Putwain https://orcid.org/0000-0001-5196-4270

Steven Howard https://orcid.org/0000-0002-1258-3210

Author note: The data reported in this study is available in a permanent online repository (https://osf.io/32av9/)

Abstract

This study sought to systematically review the full body of research on test anxiety in primary (elementary) school children aged 5–12 years. A comprehensive electronic and manual literature search identified 76 studies (85 independent samples; 53,617 children) that satisfied inclusion criteria. Inverse-variance weighted random effects meta-analysis showed that test anxiety related negatively to academic achievement, academic self-concept, and self-efficacy, and related positively to social anxiety and depression. Test anxiety was higher among girls than boys, and in Asian samples compared to European and North American samples. There was some evidence of publication bias and heterogeneity across meta-analyses. Random effects meta-regression models further showed that the association between test anxiety and mathematics achievement was stronger among older children compared to younger children, and that gender differences in test anxiety scores were more prevalent in North American samples compared to Asian samples. Intervention studies targeting anxiety reduction have been successful in reducing test anxiety and improving test anxiety-related outcomes. Overall, findings from this systematic review provide evidence that test anxiety varies in magnitude across populations and relates to multiple educational and psychosocial outcomes. We recommend further experimental studies that target the reduction of test anxiety among primary school children.

Keywords: academic performance; primary school; elementary school; test anxiety

Test anxiety in Primary School Children: A 20 year Systematic Review and Meta-Analysis

Test anxiety has been a topic of research in the education and psychology literature for over 70 years (Sarason and Mandler, 1952), it can impact on a student's performance and wellbeing as well as impact their family and school staff who show concern for the student (OECD, 2015; Lavy, 2020). Student performance on tests is a common measure of educational success in many school systems globally. Success in school then has implications for outcomes and well-being across the lifespan (OECD, 2017; Lavy, 2020). Additionally, education creates a skilled workforce leading to improved economic growth and impacts the quality of life and wellbeing of individuals (Heckman, et al., 2006; OCED, 2010).

There are many examples around the world of high stakes testing taking place in primary/elementary schools. For instance, in the US Common-Core State tests that evaluate Mathematics and ELA (English Language Arts) are used by many states in Grades 3 to 8. In England National Curriculum tests in English and Mathematics are taken at the end of Year 6 (final year of primary school). In Australia National Assessment Program for Literacy and Numeracy tests (Year 3 and Year 5) again examine Mathematics and English. In Singapore the National Primary School Leaving Examination taken at the end of primary school tests for English, Mathematics and Science. High stakes tests are tests that have real or perceived important consequences for the student (e.g., educational progression, access to job opportunities or further training, or professional accreditation; Nichols & Berliner, 2007). It has been shown that primary/elementary students perceive increased anxiety associated with high stakes standardized achievement testing (e.g., No Child Left Behind testing in the United States) compared to classroom testing (Segool et al., 2013). The results of high stakes testing can impact on a student's access to future education such as placement in schools, special programs, and training opportunities; as well as feelings of self-worth (Berliner, 2011;

Furuta et al., 2020). In terms of school administration, high stakes testing results inform school policy and funding decisions (Heissel et al., 2021). Given the perceived pressure associated with achieving high grades, it is common for children and adolescents to experience test anxiety (Lane, 2020; OCED, 2017).

Test anxiety is reported worldwide throughout many educational systems and cultures (Seipp & Schwarzer, 1996). There is now considerable international evidence that test anxiety is associated with maladaptive outcomes in secondary and tertiary education (Herzer et al., 2014; Schnell et al., 2011). However, less is known about the role of test anxiety in primary school students (Szafranski et al., 2012; von de Embse et al., 2018). This systematic review sought to synthesise protective and risk factors (correlates), experimental research, and evidence-based interventions for test anxiety in primary school students aged 5 to 12 years. Research in this age group is important as, along with physical and intellectual growth, children become more socially and emotionally aware at this age and establish educational foundations that ultimately prepare them for adolescence and adulthood (Feinstein & Bynner, 2004; Robson et al., 2020). Exploring sources of anxiety in educational contexts in this age group will help our understanding of the condition and therefore point towards future solutions.

Test Anxiety Conceptualisation

Anxiety is an unpleasant emotional state characterised by feelings of fear and distress, and is often accompanied by physiological changes (Craske & Stein, 2016). Individuals who become anxious and distressed when taking an exam are considered to be experiencing *test anxiety* – a type of performance anxiety (American Psychological Association, 2015). Test anxiety is defined as a stable, situation-specific personality trait where individuals react with elevated state anxiety to evaluative situations (Spielberger & Vagg, 1995). Trait anxiety is a stable, trait-like, disposition, thinking generally that testing situations are a threat and thus

responding with state anxiety which is a reaction to the test at the time with negative feelings such as nervousness and feeling upset (Spielberger, 1973).

Theorists originally considered test anxiety as comprising two major components: affective-physiological and cognitive (Liebert & Morris, 1967). The affective-physiological component (often termed Emotionality) involves students becoming aware of physiological arousal reactions (e.g., muscle spasms, stomach aches, nausea, sweating, and increased heart rate), which can create emotions such as fear and nervousness (Pekrun, 2006; Zeidner, 1998). The cognitive component (often termed Worry) refers to thought processes accompanying the aversive anxiety state (Spielberger & Vagg, 1995; Zeidner & Matthews, 2005). Worry can be about performance, examination conditions, the impact on the student's future, and social-evaluative concerns (Putwain, 2008). For instance, social evaluation and humiliation can be expressed through fear of judgement about exam performance by significant others (Lowe et al., 2008; Lowe & Ang, 2012).

Many theorists now consider a third component to test anxiety, namely behaviours that include ineffective study skills, avoidant or withdrawal behaviours, and procrastination (Krispenz et al., 2019; Van Eerde, 2003; Zeidner, 1998). Test anxiety behaviours can also refer to off-task behaviours occurring during an exam that indicate distraction such as tapping feet and looking around the room or at others (Wren & Benson, 2005). A high test anxious individual can be affected by different levels of these components and can experience some or all of these reactions (Zeidner, 2014). All three components of test anxiety are measured in surveys and relate to educational, psycho-educational, and psychological factors (Wren & Benson; Putwain et al., 2020).

Identified Predictors of Test Anxiety

Antecedents and correlates of test anxiety have been examined to enhance understanding of potential mechanisms and inform interventions that can target more vulnerable populations (Weems et al., 2009). Different theoretical models have also sought to explain the predictors of test anxiety with a range of psychological, psychoeducational, and demographic influences (Lowe et al., 2008; Zeidner & Matthews, 2005). The relationship between academic performance and test anxiety is the strongest finding in the literature with the landmark first study of Mandler and Sarason (1952) exploring this relationship (Hembre, 1988; von der Embse et al., 2018).

Academic Performance

Early models sought to explain the impact on test anxiety and performance. One of the earliest models, the Cognitive Attention Model (Wine, 1971), theorises that highly anxious students divide their attention between the task and self-focussing behaviours while low anxiety students focus solely on the task. In this interference model, when feeling evaluated in testing situations individuals may react with negative self-cognitions, such as lowered self-esteem, that interfere with the attention one gives to the test (Musch & Broder, 1999). The Deficit Model (Naveh-Benjamin et al., 1981; Culler & Holahan, 1980; Tobias, 1985) associates poor performance with ineffective study and test taking skills, leading to heightened emotions and high test anxiety. Building on the Cognitive Attention Model and Deficit models, the Transactional Process Model (Zeidner, 1998) considered personal factors (e.g. study skills, study habits, personality traits, self-efficacy, ability, and information processing capacity) that influence test anxiety (through their effect on Worry and Emotionality) and contribute to poor performance (Putwain & Symes, 2012; Spielberger & Vagg, 1995).

While stress around testing can motivate academic strategies and increase performance, it can also be debilitating and reduce performance (Putwain, 2007; Putwain et

al., 2012). Findings from meta-analyses suggest that high test anxiety is related to lowered performance in a number of subjects (Hembree, 1988; von der Embse et al, 2018). Test anxiety plays a role in decreasing performance through a number of theorised mechanisms that divert concentration away from the exam; for example, in testing situations maladaptive thoughts can interfere with short term memory, attention, and information processing (Eysenck et al., 2007; Liebert & Morris, 1967, Zeidner, 1998). Before the test, stress can put demands on study preparation (e.g. resulting in avoidance) and can lead to symptoms of sleeplessness and anxiety (Chamberlain et al., 2011, Spielberger & Vagg, 1995). The cognitive component of worry is more strongly related to performance than emotionality, suggesting the importance of maladaptive thoughts and rumination (Hembree, 1988; Cassady & Johnson, 2002).

Contemporary models have also moved away from an exclusive focus on performance. The Control Value Theory (Pekrun, & Perry, 2014) states that academic engagement, information processing, motivation, self-regulation, and achievement can be influenced by academic emotions such as anxiety can influence subsequent academic success and failure. Therefore, contemporary research has explored beyond the link between test anxiety and performance, looking at explanatory variables such as psychological and psychoeducational factors.

Psychological and psychoeducational factors

Psychoeducational factors can help students manage and cope with test anxiety. If the student has a positive concept of their academic ability, it can reduce the negative beliefs about testing situations (Howard, 2020; von der Embse et al., 2018). Psychoeducational measures that are negatively correlated with test anxiety include academic self-concept (a student's belief and evaluations concerning their academic skills; Marsh, 1990; Shavelson et al., 1976; Zeidner, & Schleyer, 1999), task value (motivation and belief the academic task is

valuable; Eccles et al., 1983), academic buoyancy (academic resilience to cope with academic challenges and setbacks; Martin & Marsh, 2009; Putwain & Daly, 2013), and self-efficacy (relates to beliefs in one's own organisation and ability to reach a desired outcome; Bandura, 1982; Barrow et al., 2013; Schwarzer, 1994).

Test anxiety is also associated with psychological factors and can decrease well-being through anxiety disorders, depression, and suicidal thoughts (Damer & Melendres, 2011; Putwain, 2007). Further, test anxiety is positively associated with general anxiety and other domain-specific forms of trait anxiety including social anxiety and maths anxiety (Beidel & Turner, 1988; Sarason, 1957; Zettle & Raines, 2000). General anxiety and depression are regarded as potential risk factors by predisposing children to test anxiety (Carey et al., 2017; Hocoglu et al., 2017) and are elevated in high test anxious students compared to students who report low test anxiety (Kaya, 2004; Owen et al., 2012). Both the Self-Referent Executive Processing model (S-REF: Matthews et al., 1999; Zeidner & Matthews, 2005) and Biopsychosocial Model (Lowe et al., 2008) predict that psychological factors play a role in test anxiety.

The Self-Referent Executive Processing model (S-REF: Matthews et al., 1999; Zeidner & Matthews, 2005) considers how beliefs, cognitions, and processes can initiate and maintain test anxiety. The S-REF model has three interacting levels: an executive processing system (appraisal, meta-cognitions, and approach to coping), a self-knowledge level (negative self-beliefs, avoidant motivators), and a long-term level of maladaptive interactions with the situation (hypervigilance for threat, negative feedback from others, and performance failure history). For example, in the case of high test anxiety, self-monitoring and cognitions can maintain worry through self-focused attention, negative self-appraisals, ruminative coping, and impairment of attention (Matthews et al., 1999). In the long term, negative self-knowledge can lead to greater use of avoidant and emotion-focused coping strategies, less

effective studying routines, and strengthen negative cognitions that can produce avoidant behaviours (Matthews & Wells, 2004). In contrast, the Biopsychosocial Model (Lowe et al., 2008) predicts that multiple biological, psychological, and social factors (e.g., parent/family, school influence, study habits, academic ability, and academic self-concept) play a role in test anxiety. A number of associations of test anxiety have been explored in the research to date, and by increasing statistical power previous meta-analyses have yielded significant insight in the field of test anxiety research.

Previous Meta-Analyses

Several meta-analytic reviews have synthesised research findings on test anxiety in children and adolescents. In a meta-analysis of 562 studies in child, adolescent, and adult populations, Hembree (1988) explored the nature, effects, and treatment of test anxiety, reporting *k* (number of effect sizes) and *r* (Pearson's correlation coefficient). Test anxiety in Grade 2 was unrelated to general achievement ($k = 5$, $r = –.06$) or mathematics ($k = 2$, $r = –.05$), but had a negative association with reading and literacy ($k = 5$, $r = –.15$). In older children, test anxiety was associated with general achievement in Grades 3 to post-secondary ($k = 44$, $r = –.29$), mathematics in Grades 4 to postsecondary ($k = 46$, $r = –.22$), and reading and literacy in Grades 3 to post-secondary ($k = 67$, $r = –.24$). Social anxiety in Grades 1 to 12 was also correlated with test anxiety ($k = 57$, $r = .56$). Important gender and ethnicity differences were also found. Girls showed higher levels of test anxiety than boys in Grades 1 and 2 ($k = 5$, $r = .14$), Grades 3 and 4 ($k = 26$, $r = .28$), Grades 5 to 10 ($k = 73$, $r = .43$), and Grades 11 to post-secondary ($k = 39$, $r = .27$). These effects were moderated by age such that smaller gender differences were observed in lower grades. In regard to ethnicity, higher levels of test anxiety were observed in Black compared to White students in Grades 2 to 4 ($k = 5$, $r = .52$) and in Grades 5 to 8 ($k = 8$, $r = .51$), with no ethnicity differences observed in Grades 9 to 12 ($k = 4$, $r = .02$).

The Hembree (1988) meta-analysis did not detect an effect between test anxiety and general achievement (k = 5) or mathematics (k = 2) in a Grade 2 population; this may be due to the small number of effect sizes. The larger sample size used in the present study would help to clarify the nature of this association. Hembree's (1988) meta-analysis combined studies using samples of younger and older students showing that test anxiety impacts general achievement, mathematics, reading and literacy, and social anxiety. By combining students of different ages, it is not possible to answer important questions on how test anxiety affects students in their schooling life, such as the age when test anxiety begins to impact students' performance. Additionally, the present study included outcomes not considered by Hembree (1988); psychological (general anxiety, depression, and social anxiety) and psycho-educational variables (self-efficacy and mastery goals). The present meta-analysis also incorporated a greater number of studies. For example, the Hembree (1988) analysis on gender differences included 5 studies in the lower grades whereas our meta-analysis included 18 studies in this age group.

The Richardson et al. (2012) and Seipp (1991) studies were conducted on university students and therefore do not answer specific research questions about the primary/elementary school population. The value of the present study lies in analysing primary/elementary students where high test anxiety levels are reported, and the focus on preventive and risk factors that lead to test anxiety can help to inform early intervention programs.

The von de Embse et al.'s (2018) meta-analysis showed that test anxiety is related to achievement in primary/elementary school children, but did not consider specific subject areas of mathematics, literacy, and language, alongside general school achievement. Separating the analysis into subject areas can further understanding of which curriculum areas are impacted by test anxiety. In von de Embse et al.'s (2018) study, only academic

achievement was divided into primary/elementary age groups which may limit the understanding of this specific population, as age could be a factor in the protective and risk factors for test anxiety. Additionally, von de Embse et al. (2018) did not include a moderation analysis to explain high heterogeneity that could not be accounted for by random sampling.

A subsequent meta-analysis (Seipp, 1991) synthesised findings from 126 studies on the relationship between test anxiety and academic performance on an older population, using the Achievement College Test, Stanford Achievement Test, and Grade Point Averages from schools and universities. Results showed an overall negative association of $r = –.21$ between anxiety and performance. The meta-analysis also explored separate components of test anxiety and found that Worry ($r = –.22$) showed a stronger association with academic performance than Emotionality ($r = –.15$). Effect sizes were similar for measures of state anxiety ($k = 29$, $r = –.21$) and trait anxiety ($k = 29$, $r = –.21$). The association between test anxiety and academic performance was not moderated by gender (overall standardised mean difference = .03). However, there was an important gender difference in a subsample of studies. For studies conducted in the United States ($k = 96$), girls showed higher levels of test anxiety than boys ($r = –.21$) with no significant sex differences observed in German samples or samples from other regions. In addition, a meta-analysis of psychological predictors of achievement in university students (Richardson et al., 2012) found that test anxiety related negatively to academic performance ($k = 29$, $r = –.24$). Richardson et al.'s (2012) and Seipp's (1991) studies were conducted on university students and therefore do not answer the specific research questions about the primary/elementary school population. The value of our present study lies in analysing primary/elementary students where high test anxiety levels are reported. The current research is important as it focuses on preventive and risk factors that lead to test anxiety starting and thus help to inform early intervention programs.

More recently, a comprehensive meta-analysis of 238 studies explored correlates of test anxiety in students of all ages (von de Embse et al., 2018). Test anxiety was associated with academic achievement for children in Grades 1 to 5 (k = 9, r = –.22) and Grades 6 to 8 (k = 16, r = –.25). Girls reported higher levels of test anxiety than boys in Grades 1 to 5 (k = 5, r = –.21) and Grades 6 to 8 (k = 6, r = –.16), and test anxiety was higher in ethnic minority students compared to non-ethnic minority students (k = 7, r = .12). In all age groups, test anxiety was related to self-regulation (k = 11, r = –.18), self-efficacy (k = 32, r = –.32), self-esteem (k = 5, r = –.42), extrinsic motivation (k = 5, r = .14), mastery-approach goals (k = 11, r = .05), mastery-avoidance goals (k = 7, r = .29), performance- approach goals (k = 14, r = .12), and performance-avoidance goals (k = 20, r = .31), but not locus of control (k = 6, r = .00) or intrinsic motivation (k = 13, r = –.07). Results from von de Embse et al.'s (2018) meta-analysis showed test anxiety is related to achievement in primary/elementary school children. Results, however, did not consider specific subject areas of mathematics, literacy, and language, alongside general school achievement. Therefore, separating the analysis into subject areas can assist in further understanding what curriculum areas test anxiety can impact. In von de Embse et al.'s (2018) study no other correlates besides academic achievement were divided into primary/elementary age groups which may limit the understanding of this specific population, as age could be a factor in the protective and risk factors of test anxiety. Additionally, von de Embse et al. (2018) did not include a moderation analysis to explain high heterogeneity that could not be accounted for by random sampling.

A further meta-analysis explored the relationship between test anxiety and perfectionism in all age groups (Burcas & Cretu, 2020). Of the 22 studies included, only one study sampled primary school children. The meta-analysis found that higher levels of test anxiety were related to higher levels of perfectionistic concerns (r = .28) and perfectionistic strivings (r = .19).

In addition to research syntheses of observational (cross-sectional and longitudinal) studies, several meta-analyses have also synthesised experimental research findings on test anxiety. The first meta-analysis separated interventions across treatment type (Hembree, 1988). Interventions that showed the largest effect sizes in reducing test anxiety were relaxation training in Grades 7 to postsecondary students (k = 32, d = –.68), systematic desensitisation in Grades 5 to 12 (k = 13, d = –.54), and cognitive-behavioural interventions in Grades 2 to 12 (k = 5, d = –.53). The meta-analysis included few samples of primary school students, with most studies sampling secondary school students. A subsequent meta-analysis of all age groups included 56 randomised controlled trials (RCTs) aimed at reducing test anxiety (Ergene, 2003). Results showed that it was possible to reduce test anxiety with an overall treatment effect of d = .65. The most effective programs were skill focused with behavioural (k = 45, d = .80) or cognitive (k = 16, d = .63) components.

Of further relevance to the current research synthesis, a systematic review of 10 studies explored approaches to reducing test anxiety in primary and secondary school students (von de Embse et al., 2013). Interventions using a single technique or combined approaches (multimodal interventions) were effective in reducing test anxiety. Techniques included systematic desensitisation, relaxation, academic strategy instruction, remedial academic instruction with self-instruction, guided mastery techniques, priming competence technique, biofeedback, and cognitive behavioural therapy (with and without relaxation). Findings were largely indeterminate, but techniques incorporating biofeedback, cognitive behavioural therapy, and priming competency appeared the most promising. A subsequent systematic review covering studies from 2011-2018 (Soares & Woods, 2020) updated the von de Embse et al. (2013) systematic review. Included studies were school-based interventions and focused on text anxiety reduction in children and adolescents aged 4 to 18 years. Across 11 interventions, there were mixed results with six studies showing a reduction in test

anxiety, four studies showing some success, and one study showing no reduction in test anxiety. Conclusions regarding the most successful interventions were underpinned by Cognitive Behavioural Therapy and Mindfulness psychological theories/models.

There were only a few previous meta-analyses that had an overlap of studies with the present study. The von de Embse study (2018) included 5 studies that were also in our analysis on academic achievement. The Burcas and Cretu (2020) meta-analysis on perfectionism had 22 studies included, only 1 study sampled primary school children which was included in our study. von de Embse et al. (2013) assessed interventions for test anxiety in children and adolescents from 2000-2010, our analysis included 1 study from that paper. Soares and Woods (2020) reviewed interventions for test anxiety that included all age groups from 2011-2018. There were 4 interventions on primary/elementary children out of 11 studies that we also included in our narrative review.

These systematic reviews provide evidence that test anxiety is important for multiple outcomes in older children and that interventions can be useful for reducing test anxiety. However, as far as we are aware no systematic review has exclusively synthesised this research in the primary/elementary age group. Additionally, meta-regression analysis has not been completed in previous reviews in order to explain the variation in study outcomes.

Exploratory Moderator Variables

The Biopsychosocial Model (Lowe et al., 2008) highlights that individual characteristics such as academic ability and socio-emotional functioning can influence levels of test anxiety, along with social system factors (e.g., society, community, parent/family, and school influences). Therefore, we also aim to investigate population-based differences in test anxiety scores and whether sociodemographic factors (gender, age, and world region) moderate associations between test anxiety and correlates.

Gender differences have been reported in test anxiety research, with females reporting higher test anxiety scores in all age groups (Hembree, 1988; Seipp & Schwarzer, 1996; von der Embse et al., 2018). Females could be more prone to anxiety as research indicates that they are at a greater risk of anxiety, depression, and neuroticism due to coping styles, rumination, socialisation, or genetic factors (Goodwin & Gotlib, 2004; Lim et al., 2018; Lynn & Martin, 1997; Olatunji et al., 2013, WHO, 2017). For example, females are more likely to report feeling anxiety due to the socialisation practice of showing their emotions more readily (Hill & Sarason, 1966; Kavanagh et al., 2001). Thus, tendencies towards anxiety can lead to a greater experience of test anxiety in females and is a theoretical moderator that is examined in the present review (McLean et al., 2011; Pomerantz, & Raby, 2011).

Test anxiety might differ across world regions because of varying societal factors that contribute to evaluative pressure. Higher test anxiety is reported by Asian students compared to students from other nations (Seipp & Schwarzer, 1996; Sharma & Sud, 1990; Lowe and Ang, 2012). Different values and norms in areas such as academic achievement, socialisation practices, and family expectations can play a role in test anxiety (Bodas & Ollendick, 2005). Two different hypotheses have been proposed for cultural differences in test anxiety; one relates to socialisation factors and parental expectations, and the other is the nature of high stakes exams in the education system of the country (Seipp & Schwarzer, 1996).

Age could be a potential factor contributing to test anxiety scores. Test anxiety has been found in primary/elementary students as low as 7 years of age (Connor, 2003, Klinger et al., 2015; Segool et al., 2013). Previous meta-analyses show that age does play a role in test anxiety, which increases throughout the late primary and high school years (Hembree, 1988; von der Embse et al., 2018). However, age has not been tested as a moderator in younger children, allowing exploration of, for example, whether moving into the upper years (9 to 12 years) in primary/elementary is associated with greater test anxiety.

Additionally, we pooled scores from the three most popular measures of test anxiety to give a worldwide mean that may prove beneficial as a benchmark for future studies. The three most commonly used measures of test anxiety from this review (see Table 1) are the Children's Test Anxiety Scale (CTAS; Wren & Benson, 2004), the Test Anxiety Inventory (TAI; Spielberger, 1980), and the test anxiety sub-scale of the Motivated Strategies Questionnaire (MSLQ; Pintrich et al., 1991). The CTAS, as the most popular measure in this review, served as the meta-regression model for differences in world regions in test anxiety scores.

The Current Research

The present study sought to systematically review the full body of research on test anxiety in primary/elementary school children. The review provides researchers, school psychologists, and educational staff all of the currently available evidence in the area to make informed decisions for prevention, intervention, and future research. In particular, we aim to synthesise observational studies (correlational research and longitudinal studies), experimental studies on test anxiety, and interventions for reducing test anxiety.

To date, most research on test anxiety has focused on secondary and tertiary level students (Hembree, 1988; Seipp & Schwarzer, 1996; von der Embse et al., 2018). However, as tests occur earlier in school across most world regions (von de Embse et al., 2018), a review of test anxiety research in this younger population is important. Additionally, past meta-analyses have not included moderation analysis in this age group to explore if world region, gender, age, and study quality can explain potential heterogeneity in the sample. It is important to acknowledge cultural and demographic influences in test anxiety to match with current theories (Lowe et al., 2008; Segool et al., 2014; Zeidner & Matthews, 2005).

A further purpose of the present study is to examine which individual psychological and psychoeducational factors influence test anxiety in the earlier years. Understanding the

correlates of test anxiety can be valuable to school psychologists aiming to identify ‘at risk’ populations that might benefit most from inclusion in test anxiety interventions. Systematically comparing and contrasting intervention studies can also be useful for school psychologists and teachers looking to develop effective methods of reducing test anxiety in children. Interventions were included if the aim of the study was to reduce and measure the changes in test anxiety. Therefore, understanding test anxiety from quantitative studies provides a more comprehensive picture of test anxiety in the earlier years to inform future work in this area. Therefore the aims of the research are as follows:

1. To investigate the link between test anxiety and achievement, psychoeducational, and psychological outcomes in observational research (cross-sectional and longitudinal studies).
2. To explore if test anxiety can be manipulated in experimental research in a younger age group.
3. To report on interventions which aim to reduce test anxiety in the primary/elementary student population.
4. To consider the moderators of world region, gender, age, and study quality to explain potential heterogeneity in a primary/elementary school sample.
5. To combine data from different measures of test anxiety in primary/elementary populations from around the world in order to estimate a mean score and to further explore through moderation analysis any differences in world regions in test anxiety scores.

Method

The meta-analysis was conducted in line with the PRISMA-2 2020 statement (checklist in Supplementary File S5) for the reporting of systematic reviews and meta-analyses (Page et al., 2020) and meta-analysis reporting standards from the American

Psychological Association (American Psychological Association, 2015). The full protocol was registered a priori in the international prospective register of systematic reviews (PROSPERO; CRD42020181113). All data files from quantitative results are available in a permanent online repository (https://osf.io/32av9/).

Eligibility Criteria

Observational and experimental studies of text anxiety in elementary/primary school children were eligible for inclusion. Studies were eligible for inclusion if the mean age was between 5 and 13 years. Studies that reported no mean age but had an upper age range below age 15 were also included. All measures of test anxiety were included provided the questionnaire was available and fully described. Measures of state anxiety were included provided they were completed immediately before a testing situation as such assessments are considered to be capturing test anxiety. Measures of mathematics anxiety, learning anxiety, and general school anxiety were not included. Only studies published in the English language were included. Published abstracts/conference proceedings without a full methods section were excluded.

Search Strategy

A systematic search was conducted in April 2020 using seven electronic databases. The search was restricted from January 2000 to April 2020 for a contemporary evaluation of the field. The databases used were: Scopus, Web of Science, ERIC (proquest), A+Education, PsycARTICLES, PsycINFO and MEDLINE. Three researchers developed the search terms and were based on search terms used in previous meta-analyses (e.g., Von der Embse, 2018; Jester et al., 2018). The search terms used were “test anxiety” OR “exam* anxiety” OR “exam* stress*” OR “test stress*” OR “preexam anxiety” OR “pre-exam stress” AND “elementary school*” OR “elementary educat*” OR “primary school*” OR “primary

educat*" OR "school child*" OR "children" OR "pre-adolescent child*" OR "pre-secondary" OR "young child*" OR "class*" OR "student*" OR "school*".

Study Selection

A total of 6950 records were identified through electronic databases. After title, keyword, and abstract screening the full texts of 164 studies were obtained. The first author, who has experience in systematic reviews, screened the titles, keywords, and abstracts of all studies identified through the electronic search (see Supplementary File S1 for details on search engine hits). Full texts were obtained for studies that met eligibility criteria and for studies where the relevance of the article was unclear. The first and third authors independently assessed the full texts of all identified studies using Covidence software (Babineau, 2014). The third author is an experienced anxiety theorist and has previous experience in systematic reviews. Interrater reliability between the two researchers was high, Cohen's $\kappa = .90$, $p < .001$. In instances where a discrepancy emerged between researchers regarding study inclusion, the suitability of the study was resolved through discussion between the two researchers, 8 articles were disagreed upon and discussed for inclusion; 5 were added, and 3 were excluded. The full screening procedure is illustrated in Figure 1. In total, 72 articles were included from the electronic search. A manual grey search of introduction sections and reference lists of included articles was conducted using a snowball search strategy. From this search, four additional articles were eligible for inclusion in the systematic review. In total, 76 studies were included in the systematic review.

Data Extraction and Study Quality

Data extraction was performed by the first author. The following data was extracted: author, year, the type of study design (observational/longitudinal or experimental/intervention and if study was longitudinal both waves of data were used, data was extracted from

experimental designs if they contained baseline measures of test anxiety) sample size, age range in years, mean and standard deviation of age (where no mean and standard deviation of age was provided, not provided (NP) was reported), percentage of boys/girls in the study (based on the percentage of males in the sample), and test anxiety measure. Interventions were coded for the country where the study took place, age group, number of participants, intervention sessions, control group used, mode of delivery, program supplied in the paper, outcome measures, and key results.

Coding for moderation analysis included groups of regions based on United nations definitions; regions included Europe, North America, and Asia. There was only one study each from both Africa (Ethiopia) and Oceania (Australia). For gender, the sample size and standard deviation or mean of test anxiety scores for female participants was used. The mean age of the sample was used, and if none was provided we used the median age. All studies were coded for study quality.

Study quality was assessed using the AXIS tool (Downes et al., 2016) - a critical appraisal tool for assessing research quality. The scale has 20 items that measure aspects of study quality giving a 0 or 1 score for issues such as statement of aims, detailed methods, justification of sample, statistical procedures, discussion of limitations, and reporting of conflicts of interest and ethics. A higher score denotes a higher quality study. Study quality scores were computed for all included studies (Supplementary File S3).

Meta-Analytic Strategy

In instances where data could be combined, we synthesised results using inverse variance weighted random effects meta-analysis. The inverse-variance weighted random effects model assigns a weighting to each effect size based on the inverse of its variance, giving more weight to more precise studies (Borenstein et al., 2010). The random effects model assumes within- and between-studies variability and is appropriate for combining homogeneous

studies (Overton, 1998). In instances where studies were too heterogeneous to be combined using statistical modelling, the research was synthesised using narrative analysis. For association studies, the effect size metric used was the zero-order correlation coefficient (*r*). For studies testing population-based differences, the effect size metric used was the pooled standardised mean difference (*d*). Effect sizes were taken directly from studies or were computed from means and standard deviations reported in the studies. In instances were studies reported a significant effect but did not report enough information to compute an effect size, the authors of the studies were contacted via email and the missing data were requested. In total, six authors were contacted and five provided the missing information.

Egger's regression asymmetry test was used to identify small sample effects that can indicate publication bias (Egger et al., 1997). Asymmetry in the predicted funnel shape plot alongside a significant regression coefficient ($p < .05$; two-tailed) provides evidence of publication bias. Publication bias can occur due to journal editors favouring the publication of statistically significant findings in underpowered studies meaning the final sample can be unrepresentative of the field of study (Egger et al., 1997; Rothstein et al., 2005). The potential impact of publication bias was assessed using the trim and fill procedure (Duval & Tweedie, 2000). Trim and fill uses an iterative procedure to impute additional (missing) effects on the negative side of the funnel plot, recomputing the pooled effect size at each iteration until the funnel plot is symmetric. The computed effect provides an indication of how the true effect size might appear in the absence of publication bias.

We computed the I^2 statistic as a measure of the total variation across studies due to heterogeneity rather than sampling error (Higgins et al., 2003). A general interpretation of the I^2 statistic is that a score of 25 reflects low heterogeneity, a score of 50 represents moderate heterogeneity, and a score of 75 represents high heterogeneity (Higgins & Thompson, 2002). We also report the *Q* statistic which provides a test of the hypothesis that variation in effect

sizes across studies is greater than that expected by chance alone (Huedo-Medina et al., 2006). A statistically significant *Q* statistic alongside an I^2 value above 50% should prompt the search for potential moderators of the effect (Gonzalez-Mulé & Aguinis, 2018). To test for the impact of moderating variables, we employed a protocol for random-effects meta-regression (Borenstein et al., 2009), in which the pooled mean effect size is set as the criterion variable and the moderating variable is entered as a predictor, with studies weighted by their inverse-variance weights. Meta-regression models were run for effects for which *k* was greater than 10 (Borenstein et al., 2009). Tested moderators included sample age, study quality, and world region (dummy coded as Europe, Asia, and North America). Maximum likelihood estimation method was used to calculate regression coefficients.

In addition to testing correlates and interventions of test anxiety, we also computed average scores for the most commonly used test anxiety measures to estimate normative scores across populations of the three most commonly used measures of test anxiety- CTAS (Wren & Benson, 2004), the TAI (Spielberger, 1980), and the test anxiety component of the (MSLQ; Pintrich et al., 1991). Meta-regression models were also tested for individual questionnaires to establish whether scores differ across populations. Moderators tested were age, study quality, world region, and gender. All analyses were computed using *Comprehensive Meta-Analysis 3.0* statistical software (Borenstein et al., 2014). Interpretation of effect sizes were based on current guidelines (Funder & Ozer, 2019). An effect size of r = .10 (d = .20) is considered small at the level of single events (but potentially more consequential in the long run), an effect size r = .20 (d = .41) is considered medium (some explanatory and practical use in the short run), and an effect size of r = .30 (d = .63) is considered large (impactful in the short term and potentially powerful in the long term) (Funder & Ozer, 2019).

Results

Study Characteristics

The 76 studies included 85 independent samples and 53,617 participants (grand mean age = 10.85 ± 1.35 years). The age range was between 6 and 15 years. Studies were eligible for inclusion if the mean age of participants was under 13 years old. A total of 209 effect sizes were extracted. Descriptive information on included studies is presented in Table 1. The samples were from Europe (*n* = 32), Asia (*n* = 28), North America (*n* = 18), Oceania (*n* = 1), Africa (*n* = 1), and multi-continental (*n* = 4). Based on United Nation classifications (United Nations, 2019), 64% of studies were conducted in more developed regions and 36% were conducted in less developed regions. There were 52 observational studies (50 cross sectional and 2 longitudinal) and 24 experimental studies (including 13 interventions). The mean score for study quality was 12.95 (± 2.86; range = 6 – 20) indicating that the body of research as a whole suffered from poor reporting standards.

The link between test anxiety with the variables of achievement, psychoeducational and psychological outcomes in observational research

Findings from the meta-analyses of test anxiety correlates are presented in Table 2. For achievement variables, test anxiety showed a negative association with mathematics, k = 10, r = –.21 (95% CI: –.30, –.13), literacy, k = 7, r = –.20 (95% CI: –.29, –.10), general school achievement, k = 7, r = –.23 (95% CI: –.29, –.16), but not language achievement, k = 6, r = –.11 (95% CI: –.23, .01). There was no evidence of publication bias for achievement related variables.

For psychoeducational variables, test anxiety showed a negative association with academic self-concept, k = 8, r = –.41 (95% CI: –.52, –.28). There was a significant, positive correlation between test anxiety and mathematics anxiety, k = 4, r = .57 (95% CI: .36, .73).

Test anxiety did not relate to task value, $k = 3$, $r = –.12$ (95% CI: –.54, .35), avoidance goals, $k = 2$, $r = .29$ (95% CI: –.05, .57), performance goals, $k = 2$, $r = .17$ (95% CI: –.45, .68), or mastery goals, $k = 2$, $r = –.11$ (95% CI: –.42, .22). There was no evidence of publication bias. There was substantial heterogeneity in computed effect sizes but too few effects to reliably explore moderators.

For psychological variables, test anxiety had a positive association with general anxiety, $k = 10$, $r = .62$ (95% CI: .56, .67), worry, $k = 2$, $r = .63$ (95% CI: .54, .71), physiological anxiety, $k = 2$, $r = .57$ (95% CI: .54, .61), social anxiety, $k = 3$, $r = .57$ (95% CI: .53, .60), and depression, $k = 4$, $r = .45$ (95% CI: .26, .61). Test anxiety had a negative association with self-esteem, $k = 2$, $r = –.51$ (95% CI: –.74, –.19), self-efficacy, $k = 7$, $r = –.39$ (95% CI: –.51, –.25), self-regulation, $k = 2$, $r = –.39$ (95% CI: –.61, –.11), and defensiveness, $k = 2$, $r = –.07$ (95% CI: –.12, –.02). Test anxiety was unrelated to working memory, $k = 3$, $r = –.35$ (95% CI: –.84, .14) and locus of control, $k = 2$, $r = .47$ (95% CI: –.15, .83). There was no evidence of publication bias and substantial heterogeneity for most correlates.

Manipulation of Test Anxiety in Experimental Research

Findings from the 24 experimental studies are presented in Supplementary File S3. Studies were too heterogeneous to combine using meta-analysis. There were 11 laboratory-based studies. Two of the experiments successfully manipulated state anxiety in high anxious students as measured by task accuracy in mental arithmetic, task response time, and memory recall. Four of the experiments manipulated testing situations: game-based testing compared to paper-based testing for maths; easier tests adapted to the student's ability compared to fixed or adaptive testing; high stakes testing versus low-stakes and classroom testing; and memory load in high- versus low-stress conditions. Two of the studies looked at fear appeals (i.e. fear-based messages teachers convey to students about upcoming tests) used in the

classroom prior to a test. Goals, cueing and abstract reasoning where investigated relating to test anxiety. The studies presented evidence indicating that test anxiety can be lowered through changes in personal mastery goals, high abstract reasoning skills, lowering fear appeals about tests, game-based testing, easy adaptive tests, and low stakes testing. Lowered anxiety resulted in better metacognitive skilfulness and cueing, higher scores on working memory, mental arithmetic performance, and less demand on memory recall. Most experiments showed that lowering test anxiety increased performance on a variety of tasks.

Interventions which aim to reduced test anxiety

Findings from the 13 intervention studies are presented in Supplementary File S4. Of the 13 interventions, one did not have a control group and five were non-randomised. The duration of interventions differed substantially, ranging from a single 10-minute intervention to six weeks of five sessions per week. The interventions aimed to lower test anxiety using a variety of approaches, including breathing exercises (e.g. deep breathing), writing based interventions (e.g. colouring), muscle relaxation, physical activity, test planning/dynamic training, desensitisation, yoga, assertiveness training, and cognitive behaviour therapy (CBT). Some interventions were multimodal in nature (e.g., used relaxation with exam strategies, accelerated desensitisation, or CBT). Overall, seven of the 13 studies successfully lowered test anxiety in the full sample, with a further three showing success in specific subsamples. The most successful interventions tended to include guided relaxation, breathing exercises, yoga, and CBT. Less successful interventions tended to include physical activity and writing-based tasks. Of note, three studies showed important gender differences. Interventions that appeared more successful for boys than girls included free colouring and structured breathing. Girls test anxiety was lowered successfully in an assertiveness training program targeting girls.

Moderators of world region, gender, age, and study quality

Moderators were tested for general anxiety only (k = 10). A meta-regression model showed that the positive correlation between test anxiety and general anxiety differed across continents, k = 10, $Q(2)$ = 14.66, p < .001, R^2 = .74, with a significant pairwise comparison showing that the association was stronger in European samples compared to Asian samples, b = –.30 (95% CI: –.47, – .14), p < .001. This moderation effect is depicted in Figure 2.

For gender, there was a significant difference in test anxiety scores between boys and girls, such that girls showed higher levels of test anxiety than boys, k = 18, d = .21 (95% CI: .10, .32). Effects are depicted in the Forrest plot presented in Figure 3a. However, Egger's regression asymmetry test showed some evidence of publication bias, $t(16)$ = 2.59, p = .020 (Figure 3b). Follow up trim and fill tests showed two potential missing effect sizes. Imputation of these two missing effects would reduce the effect size estimate rather substantially, pseudo d = .08 (95% CI: .03, .14). There was substantial heterogeneity allowing for exploration of potential moderators. The moderation of gender differences by world region is depicted in Figure 4. There was no significant overall effect for world region, k = 18, $Q(2)$ = 4.80, p = .091, R^2 = .22. However, a significant pairwise comparison showed that the difference in test anxiety scores between boys and girls was larger in North American samples compared to Asian samples, b = –.27 (95% CI: –.52, –.03), p = .003.

There was substantial heterogeneity allowing for exploration of moderators. For mathematics, a meta-regression model showed a significant moderation by sample age such that studies with older children showed a stronger negative association between test anxiety and mathematics performance, k = 10, b = .07 (95% CI: .02, .12), p = .007, R^2 = .64. There was no significant moderation by study quality or world region.

Measures of Test Anxiety and moderation by world region

Average scores across common test anxiety measures are presented in Table 3. Primary school children showed an average score of 62.20 (95% CI: 59.53, 64.86) on the CTAS, 39.03 (95% CI: 35.88, 42.18) on the TAI, and 2.47 (95% CI: 2.00, 2.93) on the MSLQ. These scores did not differ according to sample age or study quality. However, there was a significant moderation by world region for CTAS scores, $k = 18$, $Q(2) = 10.08$, $p = .007$, $R^2 = .39$ (Figure 5a). Observation of paired samples comparisons showed that samples in Asia scored higher on the CTAS than samples from Europe, $b = –7.67$ (95% CI: –12.70, –2.65), $p = .003$, and North America, $b = –7.14$ (95% CI: –13.01, –1.27), $p = .017$. A similar pattern of results emerged for the MSLQ, but was not significant in the smaller sample, $k = 8$, $Q(2) = 1.36$, *ns*, $R^2 = .15$ (Figure 5b). For the TAI, all samples were from Asia.

Discussion

This systematic review sought to synthesise research on test anxiety experiments, evidence-based treatments, and observational studies (correlational research and longitudinal studies) in primary/elementary school children. Previous meta-analyses of observational studies showed that test anxiety related negatively to mathematics and literacy, academic self-concept, self-efficacy, and general school achievement, and related positively to general anxiety and depression (Burcas & Cretu, 2020; Hembree, 1988; Seipp, 1991; von de Embse et al., 2018). Test anxiety was also higher among girls than boys, and in Asian samples compared to European and North American samples. Meta regression models demonstrated that the association between test anxiety and mathematics performance was stronger among older children, and that the association between trait anxiety and test anxiety was stronger in European compared to Asian samples. Meta-regression models further demonstrated that the

tendency for girls to have higher test anxiety than boys was stronger in North American samples compared to Asian samples. Narrative analyses showed that the most successful test anxiety interventions tended to include relaxation, breathing exercises, yoga, and CBT, whereas less successful interventions tended to include physical activity and writing-based tasks. Interventions that appeared more successful for boys than girls included free colouring and structured breathing (Carsley et al., 2015; Khng, 2017).

The link between test anxiety with the variables of achievement, psychoeducational and psychological outcomes in observational research

We identified 46 different variables that had been correlated with test anxiety in children. These correlates fall within four broad categories: achievement outcomes, psychoeducational factors, psychological factors, and demographic factors. In total, 23 of the 46 correlates had been tested in more than a single study and could be subjected to meta-analysis. In regard to achievement outcomes, the results showed that higher test anxiety was related to lower mathematics, literacy, and general school achievement, but was not related to language achievement. These results are generally consistent with those reported in meta-analyses of test anxiety in older students (Hembree, 1988; Richardson et al., 2012; Seipp, 1991; von der Embse et al, 2018). The finding that language performance did not significantly relate to test anxiety differs from a previous meta-analysis that found a small significant negative association of $r = –.12$ (Hembree, 1988). Interestingly, in the current meta-analysis findings showed an almost identical effect size $r = –.11$ (95% CI: –.23, .01) but with substantial heterogeneity. With only six effect sizes available, it was not possible to test moderators meaningfully and further research is needed to establish whether test anxiety might relate to language achievement in populations of young children. There was a negative association between test anxiety and mathematics was stronger among older children. This

finding is consistent with previous meta-analyses that explored different age groups and observed stronger associations in older cohorts (Barroso et al, 2020; von der Embse et al., 2018). This effect is thought to occur because of increasing pressure of schooling (Clinedinst & Koranteng, 2017; Klinger et al., 2015; Segool et al., 2013) and indicates that test anxiety becomes more relevant to performance as children get older.

For psychoeducational correlates, test anxiety showed a negative association with academic self-concept (i.e. a student's evaluation of past achievement and prospective expectations for achievement) and self-efficacy (i.e. a student's belief in their ability to successfully perform a task). A previous meta-analysis categorised self-concept under five sub-categories (locus of control, self-regulation, self-efficacy, self-esteem, and academic confidence; von de Embse, 2018) and showed that these related concepts were negatively associated with test anxiety. This finding is important as psychoeducational factors are thought to help students manage and cope with test anxiety. That is, if a student has a high self-concept of their academic ability this can act to reduce negative beliefs about themselves in testing situations (Howard, 2020; von der Embse et al., 2018).

Mathematics anxiety (defined as fear, helplessness, and worry related to mathematical performance; Ashcraft, 2002) showed a moderate positive association with test anxiety. Mathematics anxiety is a narrower construct than test anxiety and is often investigated as a separate construct using different measures (Mammarella et al., 2018). The moderate correlation observed here (r = .57) suggests that scales assessing mathematics anxiety and test anxiety are capturing distinct student experiences that are moderately related.

For psychological correlates, general measures of anxiety were moderately associated with test anxiety. In particular, children with high scores on general measures of trait anxiety (i.e., a personality disposition to experience anxiety across time and situations) as well as measures of social anxiety (i.e., persistent and intense fear of being judged or humiliated in

social situations) showed a greater tendency to experience test anxiety. This finding is consistent with previous meta-analyses in older populations (Hembree, 1988; Seipp, 1991), and could indicate that general anxiety is a risk factor that can lead to test anxiety (Carey et al., 2017; Mammarella et al., 2018).

The network theory of mental disorders (Heeren & McNally, 2016) could explain how test anxiety and other forms of anxiety are correlated. Test anxiety shares similar symptoms, such as worry and panic, with other forms of anxiety (APA, 2013) and activation of test anxiety could spread across networks of causally-related 'nodes' to render a person more vulnerable to other forms of anxiety (Putwain et al., 2021). In other words, because test anxiety and general anxiety have similar symptoms (overlapping and connected nodes), these nodes are more likely to be activated in individuals with test anxiety causing general anxiety symptomology.

Manipulation of Test Anxiety in Experimental Research

This review identified 11 laboratory-based studies that manipulated test anxiety in children. The research showed that changes in personal mastery goals, lowering fear appeals about tests, using tests with a lower cognitive load, using game-based testing (instead of paper-based), easy adaptive tests (versus a fixed test), and lower-stakes testing, all reduced test anxiety. The research also showed better metacognitive skilfulness and cueing, higher scores on working memory, mental arithmetic performance, and less demand on memory recall contributed to lowering test anxiety. Several experiments also showed that lowering test anxiety increased performance on a variety of tasks. These findings could not be synthesised using quantitative methods given that all lab-based experiments used different manipulation approaches and outcomes measures, and therefore results need to be considered as single study findings. Nevertheless, collectively they demonstrate that it is possible to

manipulate children's test anxiety in a laboratory setting using simple methods (e.g. changing the testing method to low stakes or altering messages about testing from teachers).

Anxiety manipulation can be beneficial for performance outcomes based on differing theories of test anxiety, for example, the attentional processing models would explain that more attention is on the task with lowered anxiety than on other internal emotional distractions (Eysenck et al., 2007; Wine, 1971). Further replication studies can help to pinpoint the most effective methods for manipulating anxiety in children (in terms of relative success rates) and identify an average effective size change in test anxiety scores across these methods.

Interventions which aim to reduced test anxiety

In terms of intervention studies, we identified 13 studies of which 12 included a control group and eight were randomised control trials. No harmful effects were reported in the treatment studies. Consistent with previous systematic reviews of intervention studies (Soares & Woods, 2020; von de Embse et al., 2013), findings could not be synthesised using quantitative techniques given the variety of designs and outcomes used across studies. The interventions tested including breathing exercises (e.g., diaphragmatic breathing), mindful colouring , muscle relaxation, physical activity, dynamic training, desensitisation to testing situations, yoga, assertiveness training, and CBT, with many of the interventions being multimodal. We found that seven of the 13 studies successfully lowered test anxiety, with three further studies showing success only in specific subsamples. The most successful interventions tended to include guided relaxation, breathing exercises, yoga, and CBT. This finding is consistent with previous systematic reviews that have shown CBT to be an effective method of reducing mathematics anxiety (Bicer et al., 2020) and test anxiety in older populations (Ergene, 2003; Hembree, 1988; Soares & Woods, 2020; von de Embse et al., 2013). These findings are also consistent with previous systematic reviews that have

found relaxation and breathing exercises can lower test anxiety in older samples (Hembree, 1988; Von der Embse et al., 2013). Guided relaxation, breathing exercises, and yoga are overlapping techniques that reduce parasympathetic activity associated with a threat response (Brenner et al., 2020) and our findings indicate that these techniques are useful for reducing test anxiety in younger children. CBT helps manage test anxiety through changing cognitive patterns, behaviour, and emotions by enhancing awareness of triggers and teaching skills to deal with anxiety (Chu & Harrison, 2007; Putwain & Symes, 2020).

The finding that less successful interventions tended to include physical activity and writing based tasks is notable given that physical activity (Gordon et al., 2017) and writing-based tasks (Pavlacic et al., 2019) have been shown to be effective methods of reducing anxiety in adults. These techniques might be less effective in reducing test anxiety in younger populations as children tend to be highly active without intervention (Hansen et al., 2019), rendering a physical activity intervention group similar to controls. Further, writing-based tasks might be viewed as too similar to schoolwork to actively work as a distraction from anxiety inducing stressors such as tests. Another interesting finding was that free colouring (rather than structured colouring) and structured breathing interventions appeared more successful for boys than for girls. This suggests that boys and girls tend to find different activities more relaxing resulting in different effects on test anxiety. However, these single study findings might simply reflect the rather small sample sizes included and further research is needed to establish the magnitude of this sex difference across intervention types. Perhaps the most important finding is that successful interventions (colouring in before tests, guided relaxation, breathing exercises, yoga, and CBT) are all time efficient, low cost, and easily applied to school settings before tests. A school psychologist can work with a group of children on any of these interventions, particularly the CBT, breathing exercises, or a guided meditation intervention. Additionally, school psychologists can advise teachers about

strategies to use in the classroom, such as using less fear-based messages when talking about testing or using colouring-in exercises before a test to decrease anxiety.

Moderators of world region, gender, age, and study quality

Interestingly, findings showed the association between test anxiety and general anxiety was stronger in European samples compared to Asian samples. This is a difficult moderation effect to explain and might relate to the differing levels of anxiety disorders across world regions. Anxiety disorders are more common in Europe compared to Asia (Baxter et al., 2014) and therefore are likely to play less of a role in contributing to test anxiety in Asian samples. In addition, we also found that high test anxiety scores were related to higher levels of depression. Depression has not been explored in previous meta-analyses. However, given the comorbidity of social anxiety and depression (Cummings et al., 2014; Harkin et al., 2016), this finding aligns with others and is an important area for further research. With regard to demographic factors, we found that girls reported higher levels of test anxiety than boys. This finding is consistent with previous meta-analyses in older populations (Hembree, 1988; Seipp & Schwarzer, 1996; von der Embse et al., 2018). We observed a small effect size difference (d = .21) but this accompanied evidence of publication bias with more precise studies showing smaller effects. The trim and fill correction method suggested that the true effect size might be closer to .08 (negligible) in the absence of publication bias. Importantly, high heterogeneity estimates were explained, in part, by world region. We found that the difference in test anxiety scores between boys and girls was larger in North American samples compared to Asian samples. Several hypotheses have been proposed for why girls tend to score higher on test anxiety measures. It might be the case that girls are socialised to show their emotions more readily (Hill & Sarason, 1966; Kavanagh et al., 2016) – which could help explain why we see this gender difference in North American samples and not Asian samples, where public expression of emotion is less accepted – or that

girls might feel more pressure to succeed at school than boys leading to a greater experience of test anxiety (Lofstedt et al., 2020). This finding is important for researchers and school psychologists to know that girls start to have higher test anxiety compared to boys at a younger age, and interventions can therefore be designed for this age group.

Measures of Test Anxiety and moderation by world region

Average scores across common test anxiety measures are presented in Table 3. Primary school children showed an average score of 62.20 (95% CI: 59.53, 64.86) on the CTAS, 39.03 (95% CI: 35.88, 42.18) on the TAI, and 2.47 (95% CI: 2.00, 2.93) on the MSLQ. These scores did not differ according to sample age or study quality. However, there was a significant moderation by world region for CTAS scores, $k = 18$, $Q(2) = 10.08$, $p = .007$, $R^2 = .39$ (Figure 5a). Observation of paired samples comparisons showed that samples in Asia scored higher on the CTAS than samples from Europe, $b = –7.67$ (95% CI: –12.70, –2.65), $p = .003$, and North America, $b = –7.14$ (95% CI: –13.01, –1.27), $p = .017$. A similar pattern of results emerged for the MSLQ, but was not significant in the smaller sample, $k = 8$, $Q(2) = 1.36$, *ns*, $R^2 = .15$ (Figure 5b). For the TAI, all samples were from Asia.

Regional differences in test anxiety were also found in analyses of test anxiety questionnaires. CTAS scores were higher in Asian samples compared to European and North American samples. This finding complements previous research that found small differences in test anxiety scores between individual nations (Seipp & Schwarzer, 1996). Test anxiety scores might be higher among Asian samples because of differences in parenting styles that can exert additional exam pressure on children resulting in greater test anxiety (see Spera, 2005). The biopsychosocial model (Lowe et al., 2008) also recognises the environmental influences on test anxiety including family values and socialisation factors. Cultural influence appears to have an important role in test anxiety as countries and regions often have different

values and norms surrounding academic achievement, socialisation practices, and family expectations (Bodas & Ollendick, 2005; Zeidner, 1998).

Importantly, we also found that regional differences were moderated by gender. In Asian samples, there was little difference in test anxiety scores between boys and girls with both scoring high. However, in European and North American samples gender differences were more pronounced with boys scoring lower than girls. The finding that test anxiety is higher in Asian samples appears to reflect higher test anxiety scores among boys in Asian regions. The findings that there are cultural differences in test anxiety levels in different regions of the world at such a young age is an important finding. Future research is needed to establish the environmental factors that might explain this important moderation effect. It is also important to have pooled means from all around the world for the main measures of test anxiety in primary/elementary students in order to compare population samples so that researchers and school psychologists can know if a test anxiety score is high. This measure is based on world-wide findings and incorporates different ethnicities.

Limitations

Strengths of this research synthesis include the systematic approach to study identification and the use of quantitative techniques to synthesise findings across research themes. However, there are a number of potential limitations that need to be considered in order to interpret the research findings.

First, experimental research could not be combined in a meta-analysis due to heterogeneity in study designs and outcomes tested. Therefore, only general (qualitative) conclusions can be drawn about this body of research and more research is needed before findings can be combined using statistical methods in order to identify average effect sizes and potential moderators.

A second limitation is that samples from Oceania and Africa were under-represented in the literature and whether findings generalize to these populations remains unknown.

A third limitation is that moderator effects and tests of publication bias could only be computed in instances where there were a sufficient number of effect sizes, $k > 10$ (Borenstein et al., 2009). In many instances, high heterogeneity estimates were observed and it might be the case that associations are moderated by population characteristics (e.g., gender, world region) and the relative amount of publication bias in the pooled mean effect remains unknown. Publication bias occurs because journal editors tend to favour publication of statistically significant findings in underpowered studies (Rothstein et al., 2005). Perhaps the best method to remove publication bias is for future studies to adopt registered reports (Chambers, 2013; Nosek & Lakens, 2014) as the preferred method of publication, with future meta-analyses testing effect sizes exclusively in registered report articles.

Future Research

This meta-analysis provides evidence that test anxiety is prevalent in primary/elementary school students. Further research might look to explore further the environmental factors that can lead to high levels of test anxiety in this age group (e.g., parenting). Longitudinal research that explores change in test anxiety change over several years and the factors that coincide with these changes are particularly encouraged. We also recommend further experimental research that helps to identify the most effective methods of lowering test anxiety in primary school students. CBT and breathing exercises appear particularly effective when delivered through longer programs, and more research can help pinpoint specific mechanisms that contribute to this reduction. There was also emerging evidence found indicating that certain techniques can be more effective for girls (e.g., structured colouring) and others more effective for boys (e.g., free colouring). More research is needed to establish which interventions might be more effective in particular populations

(i.e., across genders and world regions). We also encourage research into approaches that have been explored in older children.

Biofeedback and exposure have been successful in reducing test anxiety in older children (Soares & Woods, 2020), but the efficacy of these techniques in primary school children remains unknown. For interventions, we encourage researchers to adopt RCT protocols where possible in order to help establish cause-effect relationships and reduce the potential impact of confounding factors (Sibbald & Roland, 1998).

Conclusion

To conclude, this research synthesis has demonstrated that test anxiety relates negatively to mathematics and literacy achievement, academic self-concept, self-efficacy, and general school achievement, and relates positively to social anxiety and depression in primary school students. Moderation analysis showed test anxiety was higher among girls than boys, and in Asian samples compared to European and North American samples. The association between test anxiety and mathematics performance was stronger among older children compared to younger children, and gender differences in test anxiety scores were more prevalent in North American samples compared to Asian samples. Intervention studies targeting anxiety reduction appear successful in reducing test anxiety and improving test anxiety-related outcomes. These findings should be of interest to educational, school, and developmental psychologists, and school professionals interested in helping children reduce their test anxiety and perform to their potential.

School psychologists and counsellors can make a positive impact in lowering students test anxiety levels. Successful interventions to reduce test anxiety in younger children include guided relaxation, breathing exercises, colouring-in exercises, and yoga. Long-term effective interventions include CBT programs designed for test anxiety. Preventive measures will additionally be able to reduce the incidence of test anxiety – for example, advice to teachers

about providing less negative messaging around testing and encouraging positive self-efficacy around academic achievement (academic self-concept, self-efficacy, locus of control, grit, and resilience). Knowing the risk factors for test anxiety (such as general anxiety, depression, and social anxiety) can guide school psychologists to targeting these students for further assessment or intervention. Further, being aware of the higher risk of test anxiety in Asian regions, as well as in girls, especially in North American regions (compared to Europe and Asia), could be helpful to school psychologists working in these regions.

The present study explored the factors that make test anxiety better and worse in a Primary/Elementary schooling population. This knowledge helps with future targeted interventions and can be further explored in future research to help reduce the anxiety felt around testing. The findings that there are significant differences in test anxiety levels in girls and regions of the world are important findings which are found for the first time in this younger population. Having differences in world regions highlights the need for differentiated interventions based on world regions. Further research can take place on why this relationship exists which is likely to be differences in cultural practices and education systems. Thus, the data from the meta-analysis confirms the theoretical model, the Biopsychosocial Model (Lowe et al., 2008), which states the importance of social systems in the development of test anxiety. It is important for researchers and school psychologists to know that girls start to report higher test anxiety compared to boys at a younger age, this is shown to be stronger in North American samples compared to Asian samples. Interventions could therefore target girls more specifically. One intervention (an assertiveness training program) that used a girl-only population was, it was successful in reducing test anxiety (Niusha et al., 2012).

Additionally, it is important to have pooled means from all around the world for the main measures of test anxiety in primary/elementary students to compare population samples for researchers and inform school psychologists in judgements of whether test anxiety scores

are high relative to those of others This pooled measure is based on world-wide findings and incorporates different ethnicities.

We recommend further prospective studies that target anxiety change across the primary school years, and experimental research that aims to pinpoint the most effective methods of reducing test anxiety across various populations of primary school children.

References

**included in systematic review*

*Akman, B., Izgi, U., Bagce, H., & Akilli, H. I. (2007). The effect of elementary students' attitude towards science on their levels of test anxiety. *Egitim Ve Bilim-Education and Science, 32*, 3-11.

American Psychological Association (2015). *Publication manual of the American psychological association* (7th ed.). American Psychological Association.

*Arens, A. K., Becker, M., & Möller, J. (2017). Social and dimensional comparisons in math and verbal test anxiety: Within- and cross-domain relations with achievement and the mediating role of academic self-concept. *Contemporary Educational Psychology, 51,* 240–252. https://doi.org/10.1016/j.cedpsych.2017.08.005

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181-185. https://doi.org/10.1111/1467-8721.00196

Attali, Y., & Powers, D. (2010). Immediate feedback and opportunity to revise answers to open-ended questions. *Educational and Psychological Measurement, 70*(1), 22–35. https://doi.org/10.1177/0013164409332231

*Aydin, S. (2012). The effects of young EFL learners' perceptions of tests on test anxiety. *Education, 40*(2), 189-204. http://dx.doi.org/10.1080/03004279.2010.510119

*Aydin, S. (2013). Factors affecting the level of test anxiety among EFL learners at elementary schools. *International Journal of Educational Research, 4*(1), 63-81.

*Aydin, U. (2017). Test anxiety: Do gender and school-level matter? *European Journal of*

Educational Research, *6*, 187-197. https://doi.org/10.12973/eu-jer.6.2.187

*Aydin, U. (2019a). Grade level differences in the cognitive, behavioral, and physiological components of test anxiety. *International Journal of Educational Psychology*, *8*, 27-50. http://dx.doi.org/10.17583/ijep.2019.2729

*Aydin, U. (2019b). Test anxiety: Gender differences in elementary school students. *European Journal of Educational Research*, *8*, 21-30. https://doi.org/10.12973/eu-jer.8.1.21

Babineau, J. (2014). Product Review: Covidence (Systematic Review Software). *Journal of the Canadian Health Libraries Association / Journal De l'Association Des bibliothèques De La Santé Du Canada*, *35*(2), 68-71. https://doi.org/10.5596/c14-016

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, *147*(2), 134–168. https://doi.org/10.1037/bul0000307

Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., & Whiteford, H. A. (2014). The global burden of anxiety disorders in 2010. *Psychological medicine*, *44*(11), 2363-2374. https://doi.org/10.1017/s0033291713003243

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. https://doi.org/10.1080/0305764X.2011.607151

Bicer, A., Perihan, C., & Lee, Y. (2020). A Meta-Analysis: The Effects of CBT as a Clinic-& School-Based Treatment on Students' Mathematics Anxiety. *International Electronic Journal of Mathematics Education*, *15*(2), em0576. https://doi.org/10.29333/iejme/7598

*Bochis, L., & Sandra, F. (2018). Intervention program to reduce the level of test anxiety in a primary school class. A pilot study. *Revista Romaneasca Pentru Educatie Multidimensionala, 10(*4), 23-31. https://doi.org/10.18662/rrem/69

Bodas, J., & Ollendick, T. H. (2005). Test anxiety: A cross-cultural perspective. *Clinical Child and Family Psychology Review, 8*(1), 65–88. https://doi.org/10.1007/s10567-005-2342-x

*Bodas, J., Ollendick, T. H., & Sovani, A. V. (2008). Test anxiety in Indian children: A cross-cultural perspective. *Anxiety, Stress & Coping, 21,* 387-404. https://doi.org/10.1080/10615800701849902

* Boehme, K. L., Goetz, T., & Preckel, F. (2017). Is it good to value math? Investigating mothers' impact on their children's test anxiety based on control-value theory. *Contemporary Educational Psychology, 51,* 11–21. https://doi.org/10.1016/j.cedpsych.2017.05.002

*Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *Journal of Educational Research, 105*, 336-352. http://dx.doi.org/10.1080/00220671.2011.627401

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods*, *1*(2), 97-111. https://doi.org/10.1002/jrsm.12

Brenner, J., LeBlang, S., Lizotte-Waniewski, M., Schmidt, B., Espinosa, P. S., DeMets, D. L., ... & Hennekens, C. H. (2020). Mindfulness with paced breathing reduces blood pressure. *Medical hypotheses*, *142*, 109780. https://doi.org/10.1016/j.mehy.2020.109780

Burcaş, S., & Creţu, R. Z. (2020). Multidimensional perfectionism and test anxiety: A meta-analytic review of two decades of research. *Educational Psychology Review*, 1-25. https://doi.org/10.1007/s10648-020-09531-3

*Caputo, A. (2014). Psychological correlates of school bullying victimization: academic self-concept, learning motivation and test anxiety. *International Journal of Educational Psychology, 3*, 69-99. http://dx.doi.org/10.4471/ijep.2014.04

*Carey, E., Devine, A., Hill, F., & Szűcs, D. (2017). Differentiating anxiety forms and their role in academic performance from primary to secondary school. *PLoS ONE, 12*(3), e0174418. https://dx.doi.org/10.1371%2Fjournal.pone.0174418

*Carsley, D., Heath, N. L., & Fajnerova, S. (2015). Effectiveness of a classroom mindfulness coloring activity for test anxiety in children. *Journal of Applied School Psychology, 31*, 239-255. https://doi.org/10.1080/15377903.2015.1056925

*Carter, R., Williams, S., & Silverman, W. K. (2008). Cognitive and emotional facets of test anxiety in African American school children. *Cognition and Emotion, 22*(3), 539–551. https://doi.org/10.1080/02699930801886722

Cassady, J. C., & Johnson, R. E. (2002). Cognitive Test Anxiety and Academic Performance. Contemporary Educational Psychology, 27, 270-295. http://dx.doi.org/10.1006/ceps.2001.1094

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex [Editorial]. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 49*(3), 609–610.

https://doi.org/10.1016/j.cortex.2012.12.016

*Chen, Z.-H. (2014). Learning preferences and motivation of different ability students for social competition or self-competition. *Journal of Educational Technology & Society, 17*, 283-293.

Chu, B. C., & Harrison, T. L. (2007). Disorder-specific effects of CBT for anxious and depressed youth: A meta-analysis of candidate mediators of change. *Clinical Child and Family Psychology Review*, *10*(4), 352-372. https://doi.org/10.1007/s10567-007-0028-2

Clinedinst, M., & Koranteng, A. M. (2017). State of college admission.

http://hdl.handle.net/10919/83148

Craske, M. G., & Stein, M. B. (2016). Anxiety. *The Lancet, 388,* 3048–3059.

https://doi.org/10.1016/S0140-6736(16)30381-6

Culler & Holahan, 1980Benson, J., Moulin-Julian, M., Schwarzer, C., Seipp, B., &

El-Zahhar, N. (1992). Cross validation of a revised test anxiety scale using multi-national samples. In K. A. Hagvet & T. B. Johnsen (Eds.) *Advances in Test Anxiety Research Vol. 7* (pp. 62–83). Swets & Zeitlinger.

Cummings, C. M., Caporino, N. E., & Kendall, P. C. (2014). Comorbidity of anxiety and depression in children and adolescents: 20 years after. *Psychological Bulletin, 140*(3), 816–845. https://doi.org/10.1037/a0034733

*Datu, J. A. D., & Fong, R. W. (2018). Examining the association of grit with test emotions among Hong Kong Chinese primary school students. *School Psychology International, 39*, 510-525. https://doi.org/10.1177/0143034318793468

Derakhshan, N. (2020). Attentional control and cognitive biases as determinants of vulnerability and resilience in anxiety and depression. In *Cognitive Biases in Health and Psychiatric Disorders* (pp. 261-274). Academic Press.

*Donolato, E., Marci, T., Altoè, G., & Mammarella, I. C. (2020). Measuring test anxiety in primary and middle school children: Psychometric evaluation of the Test Anxiety Questionnaire for Children (TAQ-C). *European Journal of Psychological Assessment, 36*(5), 839–851. https://doi.org/10.1027/1015-5759/a000556

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal toolto assess the quality of cross-sectional studies (AXIS). *BMJ Open*, *6*(12), e011458. https://doi.org/10.1136/bmjopen-2016-011458

Driscoll, R. (2007). Westside test anxiety scale validation. ERIC Digest, ED495968.

Duval, S., & Tweedie, R. (2000). A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association, 95*(449), 89-98. https://doi.org/10.2307/2669529

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629-634. http://doi.org/10.1136/bmj.315.7109.629

Ergene, T. (2003). Effective interventions on test anxiety reduction: A meta-analysis. *School Psychology International, 24*(3), 313–328. https://doi.org/10.1177/01430343030243004

*Erturan, S., & Jansen, B. (2015). An investigation of boys' and girls' emotional experience of math, their math performance, and the relation between these variables. *European*

Journal of Psychology of Education, 30, 421-435. http://dx.doi.org/10.1007/s10212-015-0248-7

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences, 50*(7), 955–960. https://doi.org/10.1016/j.paid.2010.08.019

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Feinstein, L., & Bynner, J. (2004). The Importance of Cognitive Development in Middle Childhood for Adulthood Socioeconomic Status, Mental Health, and Problem Behavior. *Child Development, 75*(5), 1329-1339. http://www.jstor.org/stable/3696486

*Ferrando, P. J. (2015). The relevance of measuring individual discrimination in personality: A study with the Test Anxiety Scale for Children. *Personality and Individual Differences, 82*, 253-257. https://doi.org/10.1016/j.paid.2015.03.037

Friedman, I. A., & Bendas-Jacob, O. (1997). Measuring perceived test anxiety in adolescents: A selfreport scale. *Educational and Psychological Measurement, 57*(6), 1035–1046. https://doi.org/10.1177/0013164497057006012

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168.https://doi.org/10.1177/2515245919847202

Furlan, L. A., Cassady, J. C., & Pérez, E. R. (2009). Adapting the Cognitive Test Anxiety Scale for use with Argentinean university students. *International Journal of Testing, 9*(1), 3-19. https://doi.org/10.1080/15305050902733448

Furuta, J., Schofer, E., & Wick, S. (2020). The Effects of High Stakes Educational Testing on Enrollments in an Era of Hyper-Expansion: Cross-National Evidence, 1960–2010. *Social Forces*. https://doi.org/10.1093/sf/soaa071

*Gáspár, M. (2001). Interindividual stability of personality traits of children aged 9-14: A longitudinal study. *Review of Psychology, 8*(1-2), 49-59.

Goetz, T. (2004). Emotionales Erleben und selbstreguliertes Lernen bei Schülern im Fach Mathematik [Students' emotions and self-regulated learning in mathematics]. Utz.

*Goetz, T., Preckel, F., Pekrun, R., & Hall, N. C. (2007). Emotional experiences during test taking: Does cognitive ability make a difference? *Learning and Individual Differences, 17*, 3-16. https://dx.doi.org/10.1016/j.lindif.2006.12.002

Gonzalez-Mulé, E., & Aguinis, H. (2018). Advancing theory by assessing boundary conditions with metaregression: A critical review and best-practice recommendations. *Journal of Management,*

44(6), 2246–2273. https://doi.org/10.1177/0149206317710723

Gordon, B. R., McDowell, C. P., Lyons, M., & Herring, M. P. (2017). The effects of resistance exercise training on anxiety: a meta-analysis and meta-regression analysis of randomized controlled trials. *Sports Medicine*, *47*(12), 2521-2532. https://doi.org/10.1007/s40279-017-0769-0

Hankin, B.L., Snyder, H.R., Gulley, L.D., Schweizer, T.H., Bijttebier, P., Nelis, S., Toh, G., & Vasey, M.W. (2016). Understanding comorbidity among internalizing problems: Integrating latent structural models of psychopathology and risk mechanisms. *Development and Psychopathology, 28*, 987–1012. https://dx.doi.org/10.1017%2FS0954579416000663

Hansen, B. H., Kolle, E., Steene-Johannessen, J., Dalene, K. E., Ekelund, U., & Anderssen, S. A. (2019). Monitoring population levels of physical activity and sedentary time in Norway across the lifespan. *Scandinavian journal of medicine & science in sports*, *29*(1), 105-112. https://doi.org/10.1111/sms.13314

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900-1902. https://doi.org/10.1126/science.1128898

Heissel, J. A., Adam, E. K., Doleac, J. L., Figlio, D. N., & Meer, J. (2021). Testing, stress, and performance: How students respond physiologically to high-stakes testing. *Education Finance and Policy*, 16(2): 183–208. https://doi.org/10.1162/edfp_a_00306

Heeren, A., & McNally, R. J. (2018). Social anxiety disorder as a densely interconnected network of fear and avoidance for social situations. *Cognitive Therapy and Research*, *42*(1), 101-113. https://doi.org/10.1007/s10608-017-9876-3

Helmke, A. (1992). Selbstvertrauen und schulische Leistungen. Göttingen: Hogrefe.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*(1), 47–77. https://doi.org/10.2307/1170348

Herzer, F., Wendt, J. & Hamm, A.O. (2014). Discriminating clinical from nonclinical manifestations of test anxiety: A validation study. *Behavior Therapy*, *45*, 222-231. https://doi.org/10.1016/j.beth.2013.11.001

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hill, K. T., & Sarason, S. B. (1966). The relation of test anxiety and defensiveness to test and school performance over the elementary-school years: A further longitudinal study. *Monographs of the society for research in child development*, *31*(2), 1-76.

*Hocaoglu, B., & Çerkez, Y. (2017). The investigation of the relationship between perfectionism, depression, and levels of test anxiety of 4th and 5th-grade primary school students. *PONTE International Scientific Research Journal*, *73*(6), 122-133

Howard, E. (2020). A review of the literature concerning anxiety for educational assessments. *Ofqual. R*etrieved from https://www.gov.uk/government/publications/a-review-of-the-literature-onhttps://www.gov.uk/government/publications/a-review-of-the-literature-on-anxiety-for-educational-assessmentsanxiety-for-educational-assessments

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index?. *Psychological methods*, *11*(2), 193-206. https://doi.org/10.1037/1082-989x.11.2.193

*Karing, C., Doerfler, T., & Artelt, C. (2015). How accurate are teacher and parent judgements of lower secondary school children's test anxiety? *Educational Psychology, 35*, 909-925. https://doi.org/10.1080/01443410.2013.814200

Kavanagh, B. E., Ziino, S. A., & Mesagno, C. (2016). A Comparative Investigation of Test Anxiety, Coping Strategies and Perfectionism between Australian and United States Students. *North American Journal of Psychology*, *18*(3), 555-567.

*Kaya, A. (2004). Test anxiety and psychopathology in fifth-grade students in Turkey. *International Journal of Educational Reform, 13*, 19-26. https://doi.org/10.1177%2F105678790401300102

*Khng, K. H. (2017). A better state-of-mind: Deep breathing reduces state anxiety and enhances test performance through regulating test cognitions in children. *Cognition and Emotion, 31,* 1502-1510. https://doi.org/10.1080/02699931.2016.1233095

*Kiili, K., & Ketamo, H. (2018). Evaluating cognitive and affective outcomes of a digital game-based math test. *IEEE Transactions on Learning Technologies, 11*, 255-263. http://dx.doi.org/10.1109/TLT.2017.2687458

Klinger, D. A., Freeman, J. G., Bilz, L., Liiv, K., Ramelow, D., Sebok, S. S., ... & Rasmussen, M. (2015). Cross-national trends in perceived school pressure by gender and age from 1994 to 2010. *The European Journal of Public Health, 25*(2), 51-56. https://doi.org/10.1093/eurpub/ckv027

*Korhonen, J., Nyroos, M., Jonsson, B., & Eklöf, H. (2018). Additive and multiplicative effects of working memory and test anxiety on mathematics performance in grade 3 students. *Educational Psychology, 38*, 572-595. https://doi.org/10.1080/01443410.2017.1356449

Lane, S. (2020). Test-Based Accountability Systems: The Importance of Paying Attention to Consequences. *ETS Research Report Series, 2020*(1), 1-22.https://doi.org/10.1002/ets2.12283

*Larson, H. A., El Ramahi, M. K., Conn, S. R., Estes, L. A., & Ghibellini, A. B. (2010). Reducing test anxiety among third grade students through the implementation of relaxation techniques. *Journal of School Counseling, 8*, 1-19.

Lavy, S. (2020). A review of character strengths interventions in twenty-first-century schools: Their importance and how they can be fostered. *Applied Research in Quality of Life, 15*(2), 573-596. https://doi.org/10.1007/s11482-018-9700-6

*Li, H. C. W., & Chung, O. K. J. (2009). The relationship between children's locus of control and their anticipatory anxiety. *Public Health Nursing*, *26*(2), 153–160. https://doi.org/10.1111/j.1525-1446.2009.00766.x

*Lichtenfeld, S., Pekrun, R., Stupnisky, R. H., Reiss, K., & Murayama, K. (2012). Measuring students' emotions in the early years: The Achievement Emotions Questionnaire-Elementary School (AEQ-ES). *Learning and Individual Differences, 22*(2), 190–201. https://doi.org/10.1016/j.lindif.2011.04.009

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*(3, PT. 1), 975–978. https://doi.org/10.2466/pr0.1967.20.3.975

*Ling, G., Attali, Y., Finn, B., & Stone, E. A. (2017). Is a computerized adaptive test more motivating than a fixed-item test? *Applied Psychological Measurement, 41*(7), 495–511. https://doi.org/10.1177/0146621617707556

* Linnenbrink, E. A. (2005). The Dilemma of Performance-Approach Goals: The Use of Multiple Goal Contexts to Promote Students' Motivation and Learning. *Journal of Educational Psychology, 97*(2), 197–213. https://doi.org/10.1037/0022-0663.97.2.197

Löfstedt, P., García-Moya, I., Corell, M., Paniagua, C., Samdal, O., Välimaa, R., ... & Rasmussen, M. (2020). School satisfaction and school pressure in the WHO European region and North America: An analysis of time trends (2002–2018) and patterns of co-occurrence in 32 countries. *Journal of adolescent health*, *66*(6), 59-69. https://doi.org/10.1016/j.jadohealth.2020.03.007

*Lohbeck, A., Nitkowski, D., & Petermann, F. (2016). A control-value theory approach: relationships between academic self-concept, interest, and test anxiety in elementary

school children. *Child & Youth Care Forum, 45*, 887-904. http://dx.doi.org/10.1007/s10566-016-9362-1

*Lowe, P. A. (2019a). Expression and level of test anxiety in a sample of elementary students. *Higher Education Studies, 12*, 1-9. https://doi.org/10.5539/ies.v12n3p1

Lowe, P. A. (2019b). Examination of Test Anxiety in Samples of Australian and US Higher Education Students. *Higher Education Studies, 9*(4), 33-43. doi: 10.5539/hes.v9n4p33

*Lowe, P. A., Ang, R. P., & Loke, S. W. (2011a). Psychometric analyses of the Test Anxiety Scale for Elementary Students (TAS-E) scores among Singapore primary school students. *Journal of Psychopathology and Behavioral Assessment, 33*(4), 547–558. https://doi.org/10.1007/s10862-011-9250-9

*Lowe, P., & Ang, R. (2012). Cross-cultural examination of test anxiety among us and Singapore students on the Test Anxiety Scale for elementary students (TAS-E). *Educational Psychology*, *32*(1), 107–126. https://doi.org/10.1080/01443410.2011.625625

Lowe, P. A., Ang, R. P., & Loke, S. W. (2011b). Psychometric analyses of the Test Anxiety Scale for Elementary Students (TAS-E) scores among Singapore primary school students. *Journal of Psychopathology and Behavioral Assessment, 33*(4), 547–558. https://doi.org/10.1007/s10862-011-9250-9

*Lowe, P. A., Grumbein, M. J., & Raad, J. M. (2011). Examination of the psychometric properties of the test anxiety scale for elementary students (TAS-E) scores. *Journal of Psychoeducational Assessment, 29*, 503-514. https://doi.org/10.1177%2F0734282910395894

Lowe, P. A., Lee, S. W., Witteborg, K. M., Prichard, K. W., Luhr, M. E., Cullinan, C. M., Mildren, B. A., Raad, J. M., Cornelius, R. A., & Janik, M. (2008). The Test Anxiety Inventory for Children and Adolescents (TAICA): Examination of the psychometric properties of a new multidimensional measure of test anxiety among elementary and secondary school students. *Journal of Psychoeducational Assessment, 26*(3), 215–230. https://doi.org/10.1177/0734282907303760

Mammarella, I. C., Donolato, E., Caviola, S., & Giofrè, D. (2018). Anxiety profiles and protective factors: A latent profile analysis in children. *Personality and Individual Differences, 124,* 201–208. https://doi.org/10.1016/j.paid.2017.12.017

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology, 21*(1), 89–101. https://doi.org/10.1080/01443410020019867

*Mammarella, I. C., Donolato, E., Caviola, S., & Giofrè, D. (2018). Anxiety profiles and protective factors: A latent profile analysis in children. *Personality and Individual Differences, 124,* 201–208. https://doi.org/10.1016/j.paid.2017.12.017

Matthews, G., & Wells, A. (2004). Rumination, depression, and metacognition: The S-REF model. *In* Papageorgiou, C., & Wells, A. (Eds.). (2004). *Depressive rumination: Nature, Theory and Treatment*. John Wiley & Sons.

Matthews, G., Hillyard, E. J., & Campbell, S. E. (1999). Metacognition and maladaptive coping as components of test anxiety. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice, 6*(2), 111-125. https://doi.org/10.1002/(SICI)1099-0879(199905)6:2<111::AID-CPP192>3.0.CO;2-4

*Mavilidi, M. F., Hoogerheide, V., & Paas, F. (2014). A quick and easy strategy to reduce test anxiety and enhance test performance. *Applied Cognitive Psychology, 28,* 720-726. https://doi.org/10.1002/acp.3058

*Mavilidi, M. F., Ouwehand, K., Riley, N., Chandler, P., & Paas, F. (2020). Effects of an Acute Physical Activity Break on Test Anxiety and Math Test Performance. *International journal of environmental research and public health*, *17*(5), 1523. https://doi.org/10.3390/ijerph17051523

McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *Journal of psychiatric research*, *45*(8), 1027-1035. https://doi.org/10.1016/j.jpsychires.2011.03.006

*Metallidou, P., & Vlachou, A. (2007). Motivational beliefs, cognitive engagement, and achievement in language and mathematics in elementary school children. *International Journal of Psychology*, *42*(1), 2–15. https://doi.org/10.1080/00207590500411179

*Miller, M., Morton, J., Driscoll, R., & Davis, K. A. (2005). Accelerated desensitization with adaptive attitudes and test gains with 5th graders. http://ezproxy.uow.edu.au/login?url=https://search.proquest.com/docview/62013992?accountid=15112

*Miller, N., DeLapp, R., & Driscoll, R. (2007). Group Anxiety Reduction with Sixth Grade Students.http://ezproxy.uow.edu.au/login?url=https://search.proquest.com/docview/62057632?accountid=15112

*Mulvenon, S. W., Stegman, C. E., & Ritter, G. (2005). Test Anxiety: A Multifaceted Study on the Perceptions of Teachers, Principals, Counselors, Students, and Parents. *International Journal of Testing*, *5*(1), 37–61. https://doi.org/10.1207/s15327574ijt0501_4

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.

*Narendra, B. M., Rai, S. C., & Rashmitha, A. P. (2019). Effect of yoga on anxiety in school going children of age 12-14 years. *Indian Journal of Public Health Research and Development, 10,*

439-444. https://doi.org/10.37506/ijphrd.v10i8.6910

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y.-g., & Holinger, D. P. (1981). Test anxiety: Deficits in information processing. *Journal of Educational Psychology, 73*(6), 816–824. https://doi.org/10.1037/0022-0663.73.6.816

*Ng, E., & Lee, K. (2010). Children's task performance under stress and non-stress conditions: A test of the processing efficiency theory. *Cognition and Emotion, 24*(7), 1229–1238. https://doi.org/10.1080/02699930903172328

*Ng, E., & Lee, K. (2015). Effects of trait test anxiety and state anxiety on children's working memory task performance. *Learning and Individual Differences, 40,* 141-148. https://doi.org/10.1016/j.lindif.2015.04.007

*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

*Niusha, B., Farghadani, A., & Safari, N. (2012). *Effects of assertiveness training on test anxiety of girl students in first grade of guidance school*. In G. A. Baskan, F. Ozdamli, S. Kanbul, & D. Ozcan (Eds.), 4th World Conference on Educational Sciences (Vol. 46, pp. 1385-1389). https://dx.doi.org/10.5812%2Fircmj.21096

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*(3), 137-141. http://dx.doi.org/10.1027/1864-9335/a000192

*Nyroos, M., & Wiklund-Hörnqvist, C. (2011). Introducing national tests in Swedish primary education: Implications for test anxiety. *Electronic Journal of Research in Educational Psychology, 9*(3), 995–1022.

*Nyroos, M., Jonsson, B., Korhonen, J., & Eklöf, H. (2015a). Children's mathematical achievement and how it relates to working memory, test anxiety and self-regulation: A person-centred approach. *Education Inquiry, 6*(1). https://doi.org/10.3402/edui.v6.26026

*Nyroos, M., Korhonen, J., Peng, A., Linnanmäki, K., Svens-Liavåg, C., Bagger, A., & Sjöberg, G. (2015b). Cultural and gender differences in experiences and expression of test anxiety among Chinese, Finnish, and Swedish grade 3 pupils. *International Journal of School & Educational Psychology, 3*(1), 37–48. https://doi.org/10.1080/21683603.2014.915773

OECD (2010), *The High Cost of Low Educational Performance: The Long-run Economic Impact of Improving PISA Outcomes*, PISA, OECD Publishing. Retrieved from https://doi.org/10.1787/9789264077485-en.

Organisation for Economic Co-operation and Development (OECD). (2017). *PISA 2015 results (Volume III): Students' Well-Being*. OECD. OECD Publishing. Retrieved from http://www.oecd.org/education/pisa-2015-results-volume-iii-9789264273856-en.htm

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3*(3), 354–379. https://doi.org/10.1037/1082-989X.3.3.354

*Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. *School Psychology International*, *33*(4), 433–449. https://doi.org/10.1177/0143034311427433

Page, M. J., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., mulrow, c. d., … Moher, D. (2020, September 14). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. https://doi.org/10.31222/osf.io/v7gm2

Pavlacic, J. M., Buchanan, E. M., Maxwell, N. P., Hopke, T. G., & Schulenberg, S. E. (2019). A meta-analysis of expressive writing on posttraumatic stress, posttraumatic growth, and quality of life. *Review of General Psychology*, *23*(2), 230–250. https://doi.org/10.1177/1089268019831645

Pekrun, R., & Perry, R. P. (2014). *Control-value theory of achievement emotions.* In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *Educational psychology handbook series. International handbook of emotions in education* (p. 120–141). Routledge/Taylor & Francis Group.

*Peleg-Popko, O., & Klingman, A. (2002). Family environment, discrepancies between perceived actual and desirable environment, and children's test and trait anxiety. *British Journal of Guidance & Counselling*, *30*, 451-466. https://doi.org/10.1080/0306988021000025646

Pintrich, P., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ).* University of Michigan.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

Pomerantz, S., & Raby, R. (2011). 'Oh, she's so smart': girls' complex engagements with post/feminist narratives of academic success. *Gender and education, 23*(5), 549-564. 10.1080/09540253.2010.538014

Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional & Behavioural Difficulties, 13*(2), 141–155. https://doi.org/10.1080/13632750802027713

*Putwain, D. W., & Best, N. (2011). Fear appeals in the primary classroom: Effects on test anxiety and test grade. *Learning and Individual Differences, 21*(5), 580–584. https://doi.org/10.1016/j.lindif.2011.07.007

*Putwain, D. W., & Best, N. (2012). Do highly test anxious students respond differentially to fear appeals made prior to a test? *Research in Education, 88,* 1-10. http://dx.doi.org/10.7227/RIE.88.1.1

Putwain, D.W., & Symes, W. (2020). *The Four Ws of Test Anxiety. Psychologica*, 63(2), 31-52. https://doi.org/10.14195/1647-8606_63-2_2

Putwain, D. W., & Symes, W. (2012). Achievement goals as mediators of the relationship between competence beliefs and test anxiety. *British Journal of Educational Psychology, 82*(2), 207–224. https://doi.org/10.1111/j.2044-8279.2011.02021.x

*Putwain, D. W., Nicholson, L. J., Connors, L., & Woods, K. (2013). Resilient children are less test anxious and perform better in tests at the end of primary schooling. *Learning and Individual Differences, 28,* 41–46. https://doi.org/10.1016/j.lindif.2013.09.010

Putwain, D. W., & von der Embse, N. P. (2021). Cognitive–behavioral intervention for test anxiety in adolescent students: do benefits extend to school-related wellbeing and clinical anxiety. *Anxiety, Stress, & Coping*, *34*(1), 22-36. https://doi.org/10.1080/10615806.2020.1800656

Putwain, D. W., von der Embse, N. P., Rainbird, E. C., & West, G. (2020). The development and validation of a new Multidimensional Test Anxiety Scale (MTAS). *European Journal of Psychological Assessment.* Advance online publication. https://doi.org/10.1027/1015-5759/a000604

*Raju, P. M., & Asfaw, A. (2009). Recalled test anxiety in relation to achievement, in the context of general academic self-concept, study habits, parental involvement and socio-economic status among Grade 6 Ethiopian students. *Education 3-13, 37*(3), 269–285. https://doi.org/10.1080/03004270902734085

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, *146*(4), 324–354. https://doi.org/10.1037/bul0000227

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis.

Publication bias in meta-analysis: Prevention, assessment and adjustments. Wiley.

*Rouxel, G. (2001). Cognitive-affective determinants of performance in mathematics and verbal domains - Gender differences. *Learning and Individual Differences, 12*, 287-310. https://doi.org/10.1016/S1041-6080(01)00041-3

Sarason, S. B., & Mandler, G. (1952). Some correlates of test anxiety. *The Journal of Abnormal and Social Psychology*, *47*(4), 810–817. https://doi.org/10.1037/h0060009

Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R., & Ruebush, B. K. (1960). *Anxiety in Elementary School Children*. Wiley & Sons Inc. https://doi.org/10.1037/14349-000

*Schnell, K., Tibubos, A. N., Rohrmann, S., & Hodapp, V. (2013). Test and math anxiety: A validation of the German Test Anxiety Questionnaire. *Polish Psychological Bulletin, 44*(2), 193–200. https://doi.org/10.2478/ppb-2013-0022

Schnell, K., Tibubos, A., Rohrmann, S., Schienle, A., & Hodapp, V. (2011). Test anxiety in highschool students: Self-reports and students-teacher agreement. *Stress and Anxiety: Application to Education and Health*, 45-54. Buchwald, P., Moore, K. A., & Ringeisen, T. (Eds.). (2011). *Stress and anxiety: application to education and health*. Logos Verlag.

*Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened test anxiety among young children: Elementary school students' anxious responses to high-stakes testing. *Psychology in the Schools, 50*(5), 489–499. https://doi.org/10.1002/pits.21689

Segool, N. K., von der Embse, N. P., Mata, A. D., & Gallant, J. (2014). Cognitive behavioral model of test anxiety in a high-stakes context: An exploratory study. *School Mental Health: A Multidisciplinary Research and Practice Journal, 6*(1), 50–61. https://doi.org/10.1007/s12310-013-9111-7

Seipp, B., 1991. Anxiety and academic performance: a meta-analysis of findings. *Anxiety Research*, *4*(1), 27–41. https://doi.org/10.1080/08917779108248762

*Selkirk, L. C., Bouchey, H. A., & Eccles, J. S. (2011). Interactions among domain-specific expectancies, values, and gender: Predictors of test anxiety during early adolescence. *The Journal of Early Adolescence*, *31*(3), 361–389. https://doi.org/10.1177/0272431610363156

Seipp, B., & Schwarzer, C. (1996). Cross-cultural anxiety research. A review. In C. Schwarzer & M. Zeidner (Eds.), *Stress, anxiety, and coping in academic settings* (pp. 13–68). Tubingen: Francke. https://doi.org/10.1007/s10567-005-2342-x

Sibbald, B., & Roland, M. (1998). Understanding controlled trials. Why are randomised controlled trials important?. *BMJ: British Medical Journal*, *316*(7126), 201. https://doi.org/10.1136/bmj.316.7126.201

Spera, C. (2005). A Review of the Relationship Among Parenting Practices, Parenting Styles, and Adolescent School Achievement. *Educational Psychology Review, 17*(2), 125–146. https://doi.org/10.1007/s10648-005-3950-1

*Shih, S.-S. (2005). Role of Achievement Goals in Children's Learning in Taiwan. *The Journal of Educational Research, 98*(5), 310–319. https://doi.org/10.3200/JOER.98.5.310-319

*Shoahosseini, R., & Baghaei, P. (2020). Validation of the persian translation of the children's test anxiety scale: a multidimensional rasch model analysis. *European Journal of Investigation in Health, Psychology and Education*, *10*, 59-69. https://doi.org/10.1177/2158244014555113

*Shores, M. L., Shannon, D. M., & Smith, T. G. (2010). Individual learner variables and their effect on mathematics achievement as students advance from fifth to sixth grade. *Journal of Research in Childhood Education, 24*(3), 187–194. https://doi.org/10.1080/02568543.2010.487393

Smits, J. A. E., & Vorst, H. C. M. (2013). *School questionnaire.* Pearson.

Soares, D., & Woods, K. (2020). An international systematic literature review of test anxiety interventions 2011–2018. *Pastoral Care in Education*, *38*(4), 311-334.

https://doi.org/10.1080/02643944.2020.1725909

Spera, C. (2005). A Review of the Relationship Among Parenting Practices, Parenting Styles, and Adolescent School Achievement. *Educational Psychology Review, 17*(2), 125–146. https://doi.org/10.1007/s10648-005-3950-1

Spielberger, C. D. (1980). *Preliminary professional manual for the test anxiety inventory*. Consulting Psychologists Press.

Spielberger, C. D. (1973). *Manual for the state-trait anxiety inventory for children*. Consulting Psychologists Press.

Spielberger, C. D., & Vagg. R.P. (1995). Test Anxiety: A Transactional Process Model. In by C. D.

Speilberger & P. R. Vagg (Eds.) *Test Anxiety: Theory, Assessment and Treatment*, (pp. 3–14). Taylor & Francis.

Steinmayr, R., Crede, J., McElvany, N., & Wirthwein, L. (2016). Subjective well-being, test anxiety, academic achievement: Testing for reciprocal effects. *Frontiers in Psychology, 6*(1994). *https://doi.org/10.3389/fpsyg.2015.01994*

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1

Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test anxiety inventory: 30 years later. *Anxiety, Stress & Coping, 25*(6), 667-677. https://doi.org/10.1080/10615806.2012.663490

*Thompson, H. R., Duvall, J., Padrez, R., Rosekrans, N., & Madsen, K. A. (2016). The impact of moderate-vigorous intensity physical education class immediately prior to standardized testing on student test-taking behaviors. *Mental Health and Physical Activity, 11,* 7-12. http://dx.doi.org/10.1016%2Fj.mhpa.2016.06.002

Tobias.S (1985) Test Anxiety: Interference, Defective Skills, and Cognitive Capacity, *Educational Psychologist, 20*(3), 135-142. https://doi.org/10.1207/s15326985ep2003_3

Turner, B. G., Beidel, D. C., Hughes, S., & Turner, M. W. (1993). Text anxiety in African American school children. *School Psychology Quarterly, 8*(2), 140–152. https://doi.org/10.1037/h0088835

*Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress & Coping: An International Journal, 13*(4), 391–412. https://doi.org/10.1080/10615800008248343

*Villares, E., Colvin, K., Carey, J., Webb, L., Brigman, G., & Harrington, K. (2014). Convergent and Divergent Validity of the Student Engagement in School Success Skills Survey. *Professional Counselor, 4(5),* 541-552. doi:10.15241/ev.4.5.541

*Vogelaar, B., Bakker, M., Elliott, J. G., & Resing, W. C. (2017). Dynamic testing and test anxiety amongst gifted and average-ability children. *British Journal of Educational Psychology, 87,* 75-89. https://doi.org/10.1111/bjep.12136

von der Embse, N., Barterian, J., & Segool, N. (2013). Test anxiety interventions for children and adolescents: A systematic review of treatment wrentudies from 2000–2010. *Psychology in the Schools, 50*(1), 57-71. https://doi.org/10.1002/pits.21660

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.doi: 10.1016/j.jad.2017.11.048

Wieczerkowski, W., Nickel, H., Janowski, A., Fittkau, B., & Rauer, W. (1981). Angstfragebogen fu¨r Schu¨ ler (AFS) [Anxiety questionnaire for students]. Hogrefe.

Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology, 80*(2), 210–216. https://doi.org/10.1037/0022-0663.80.2.210

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

*Wood, S. G., Hart, S. A., Little, C. W., & Phillips, B. M. (2016). Test anxiety and a high-stakes standardized reading comprehension test: a behavioral genetics perspective. *Journal of*

Developmental Psychology, 62, 233-251. https://doi.org/10.13110/merrpalmquar1982.62.3.0233 *Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress & Coping: An International Journal, 17*(3), 227–240. https://doi.org/10.1080/10615800412331292606

*Yeo, L. S., Goh, V. G., & Liem, G. A. D. (2016). School-based intervention for test anxiety. *Child & Youth Care Forum, 45*(1), 1–17. https://doi.org/10.1007/s10566-015-9314-1

Zeidner, M. (1998). *Test Anxiety. The State of the Art*, Kluwer Academic Publishers.

Zeidner, M. (2014). *Anxiety in education.* In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *Educational psychology handbook series. International handbook of emotions in education* (p. 265–288). Routledge/Taylor & Francis Group.

Zeidner, M. (2020). Test anxiety. *The Wiley Encyclopedia of Personality and Individual Differences: Personality Processes and Individual Differences,* (pp. 445-449). Wiley.

Zeidner, M., Nevo, B., & Lipschitz, H. (1988). The Hebrew version of the test anxiety inventory. *Haifa: University of Haifa*.

Zeidner, M., & Matthews, G. (2005). Evaluation anxiety. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 141–163). Guilford Press.

*Zhang, Y., Watermann, R., & Daniel, A. (2016). Are multiple goals in elementary students beneficial for their school achievement? A latent class analysis. *Learning and Individual Differences, 51,* 100–110. https://doi.org/10.1016/j.lindif.2016.08.023

*Zhou, J., & Urhahne, D. (2013). Teacher judgment, student motivation, and the mediating effect of attributions. *European Journal of Psychology of Education, 28*(2), 275–295. https://doi.org/10.1007/s10212-012-0114-9

*Zhu, M., & Urhahne, D. (2014). Assessing teachers' judgements of students' academic motivation and emotions across two rating methods. *Educational Research and Evaluation, 20*(5), 411–427. https://doi.org/10.1080/13803611.2014.964261

Table 1

Descriptive information on included studies

Research characteristics	Number of Studies (%)
Research Design	
Observational	**52 (68%)**
Longitudinal	2 (3%)
Cross-sectional	50 (66%)
Experimental	**24 (32%)**
Interventions	13 (17%)
Laboratory based studies	11 (15%)
Measures	
CTAS	33 (35%)
TAI	16 (15%)
STAIC-S	10 (11%)
MSLQ	9 (10%)
TAS-C	6 (6%)
WTAS	5 (5%)
TAS-E	4 (4%)
FTA	3 (3%)
AEQ	3 (3%)
TAQ-C	3 (3%)
World Regions	
Europe	32 (40%)
Asia	28 (35%)
North America	18 (23%)
Oceania	1 (< 1%)
Africa	1 (< 1%)
Latin America	0
Development region	
More developed region	51 (64%)
Less developed region	29 (36%)

Note: AEQ = Achievement Emotions Questionnaire (Pekrun et al., 2005); CTAS = Child Test Anxiety Scale (Wren & Benson, 2004); FTA = FRIEDBEN Test Anxiety Scale for adolescence (Friedman & Bendas-Jacob, 1997); MSLQ = Motivated Strategies for Learning Questionnaire (Pintrich et al., 1991, 1993); STAIC-S = State-Trait Anxiety Inventory for Children (Spielberger, 1973); State anxiety/emotions before and after exam (Goetz, 2004); TAI = Test anxiety inventory (Spielberger, 1980); TAQ-C = Test Anxiety Questionnaire for Children (Donolato et al., in press); TASC = Test Anxiety Scale for Children (Sarason et al., 1960);TAS-E = Test Anxiety Scale for Elementary Students (Lowe et al., 2011); WTAS = The Westside Test Anxiety Scale (Driscoll, 2007).

Table 2

Random effects mean associations for test anxiety as it relates to components of achievement, psychoeducational and psychological correlates, and demographic factors

Measure	*k*	*n*	*r* [95% CI]	I^2	*Q*	*E(p)*
Achievement						
Mathematics performance	10	13,169	-.21 [-.30, -.13]	94.53	164.55**	.90
Literacy performance	7	2627	-.20 [-.29, -.10]	77.09	26.19**	.25
Language performance	6	10,117	-.11 [-.23, .01]	95.32	107.34**	.55
General school achievement	7	5741	-.23 [-.29, -.16]	57.69	14.18*	.43
Psychoeducational correlates						
Academic self-concept	8	7449	-.41 [-.52, -.28]	96.52	205.38**	.25
Self-efficacy	7	6836	-.39 [-.51, -.25]	96.23	159.15**	.90
Maths anxiety	4	2463	.57 [.36, .73]	97.70	130.17**	.68
Task value	3	1038	-.12 [-.54, .35]	98.13	107.19**	.18
Performance-Avoidance goals	2	4585	.29 [-.05, .57]	95.73	23.39**	N/A
Performance-Approach goals	2	4585	.17 [-.45, .68]	98.80	83.55**	N/A
Mastery goals	2	4585	-.11 [-.42, .22]	95.30	21.29**	N/A
Expectations of teachers/parents	1	430	-.63 [-.57, -.69]	0.00	0.00	N/A
Expectations of self	1	430	-.63 [-.57, -.69]	0.00	0.00	N/A

Shame	1	1051	.58 [.54, .62]	0.00	0.00	N/A
Hope	1	1051	-.31 [-.36, -.25]	0.00	0.00	N/A
Pride	1	1051	-.15 [-.21, -.09]	0.00	0.00	N/A
Expectancy for success	1	272	-.32 [-.42, -.21]	0.00	0.00	N/A
Engagement	1	789	-.32 [-.38, -.26]	0.00	0.00	N/A
Study habits	1	497	-.25 [-.33, -.17]	0.00	0.00	N/A
Cognitive strategies	1	4342	-.20 [-.23, -.17]	0.00	0.00	N/A
Academic buoyancy	1	664	-.15 [-.22, -.08]	0.00	0.00	N/A
Academic interest	1	192	-.05 [-.19, .09]	0.00	0.00	N/A
Intrinsic motivation	1	4387	.02 [-.01, .05]	0.00	0.00	N/A
Self-direction of learning	1	4342	-.02 [-.05, .01]	0.00	0.00	N/A
Support of classmate learning	1	4342	-.02 [-.05, .01]	0.00	0.00	N/A
Psychological correlates						
General anxiety	10	4670	.62 [.56, .67]	89.19	83.25**	.62
Depression	4	784	.45 [.26, .61]	85.72	21.00**	.27
Social anxiety	3	1560	.57 [.53, .60]	0.00	1.41**	.28
Working memory	3	152	-.35 [-.84, .14]	43.15	3.52	.84
Worry (RCMAS-2)	2	1400	.63 [.54, .71]	84.05	6.27*	N/A
Physiological anxiety (RCMAS-2)	2	1400	.57 [.54, .61]	0.00	0.00	N/A

Defensiveness (RCMAS-2)	2	1400	-.07 [-.12, -.02]	0.00	0.55**	N/A
Self-esteem	2	1209	-.51 [-.74, -.19]	97.58	41.30**	N/A
Locus of control	2	982	.47 [-.15, .83]	99.10	111.23	N/A
Self-regulation	2	4966	-.39 [-.61, -.11]	98.05	51.31**	N/A
Resilience	1	123	-.47 [-.60, -.32]	0.00	0.00	N/A
Grit (consistency of interests)	1	1051	.35 [.30, .40]	0.00	0.00	N/A
Grit (perseverance of effort)	1	1051	-.06 [-.12, .00]	0.00	0.00	N/A
Perfectionism (social-oriented)	1	210	.28 [.15, .40]	0.00	0.00	N/A
Perfectionism (self-oriented)	1	210	.19 [.06, .32]	0.00	0.00	N/A
Parent involvement	1	497	-.21 [-.29, -.12]	0.00	0.00	N/A
Ego resiliency	1	664	-.08 [-.15, -.00]	0.00	0.00	N/A
Self-regulation of arousal	1	4342	.08 [.05, .11]	0.00	0.00	N/A
Demographic factors						
Education	2	487	.06 [-.19, .31]	85.92	7.10*	N/A
Socioeconomic status	1	497	-.14 [-.23, -.05]	0.00	0.00	N/A

Note. k = number of pooled effect sizes; n = number of pooled participants; r = mean effect size expressed as Pearson correlation; *CI* = confidence interval; I^2 = heterogeneity estimate expressed as a percentage; Q = heterogeneity estimate; E(p) = significance value for Egger's regression asymmetry test- two tailed (values below .05 indicate meaningful publication bias); N/A- publication bias not applicable to less than 3 pooled effect sizes.

*p < .05, **p < .01

Table 3

Meta-analysis and meta-regression models for moderators of test anxiety measures

Measure	Meta-analysis				Meta-regression					
	k	*n*	Mean	95% CI	Age		Study Quality		Region	
					b(s.e.)	R^2	*b*(s.e.)	R^2	*Q*(df)	R^2
Test anxiety measures										
CTAS	18	7199	62.20	59.53, 64.86	.42 (-1.96, 2.79)	.01	-.01 (-1.17, 1.13)	.00	10.08(2)**	.39
TAI	12	779	39.03	35.88, 42.18	-.86 (-2.61, 0.89)	.08	.57 (-.12, 1.27)	.18	-	-
MSLQ	8	9070	2.47	2.00, 2.93	-.12 (-.67, .43)	.02	-.05 (-.32, .22)	.02	1.36(2)	.15

Note: Random effects reported. CTAS = Child Test Anxiety Scale; TAI = Test Anxiety Inventory (for the TAI all samples were from Asia); MSLQ = Motivated Strategies Questionnaire.

**$p < .01$

Figure 1

Flow diagram of the screening process.

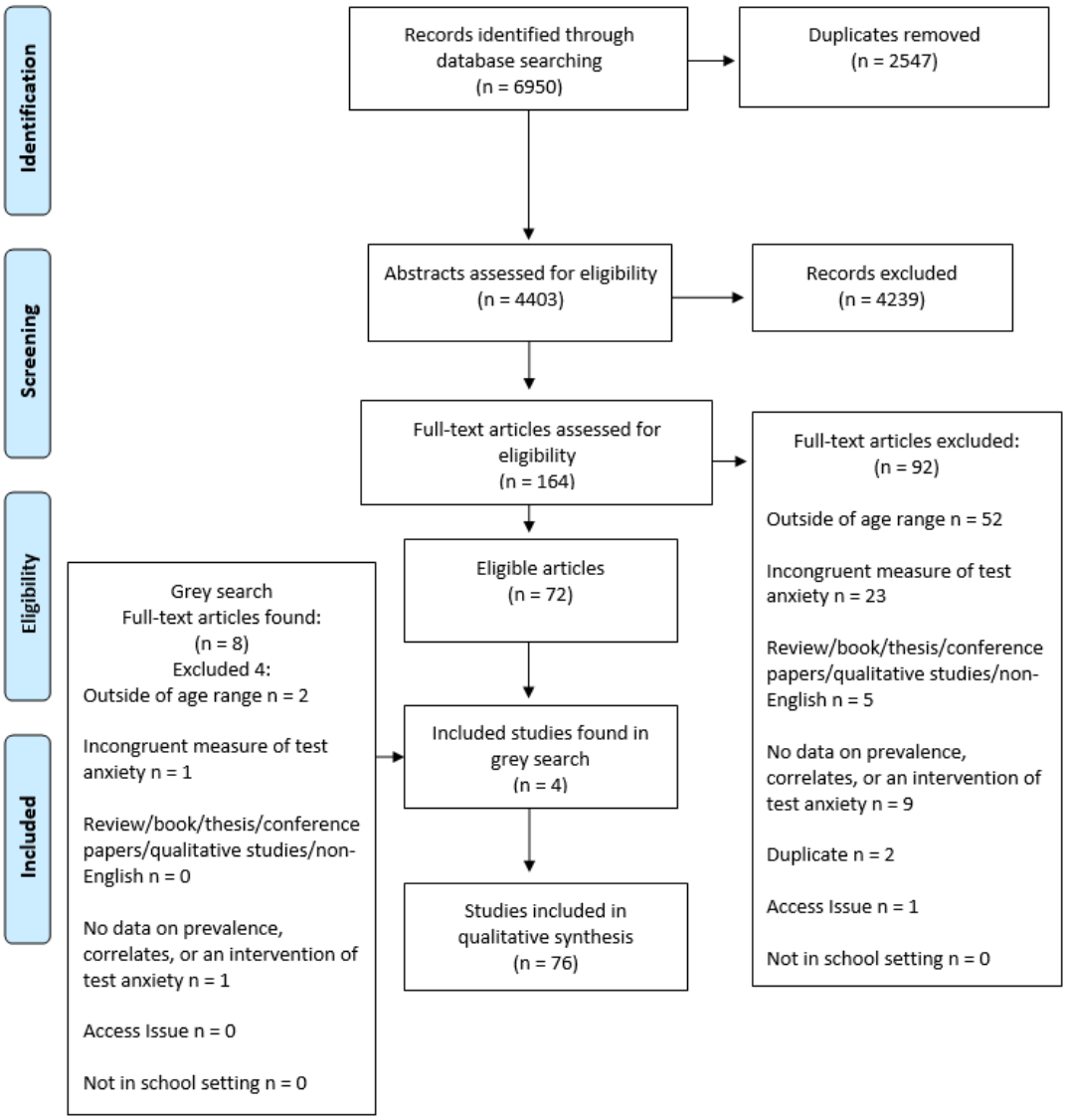


Figure 2

Graphical representation of moderation analysis for the association between test anxiety and general anxiety.

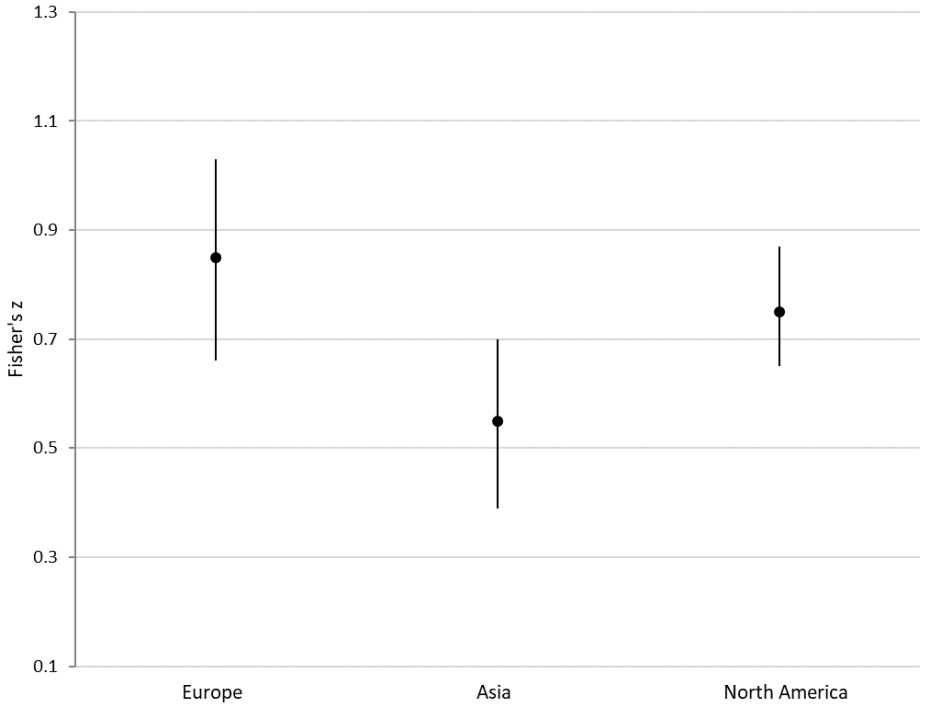
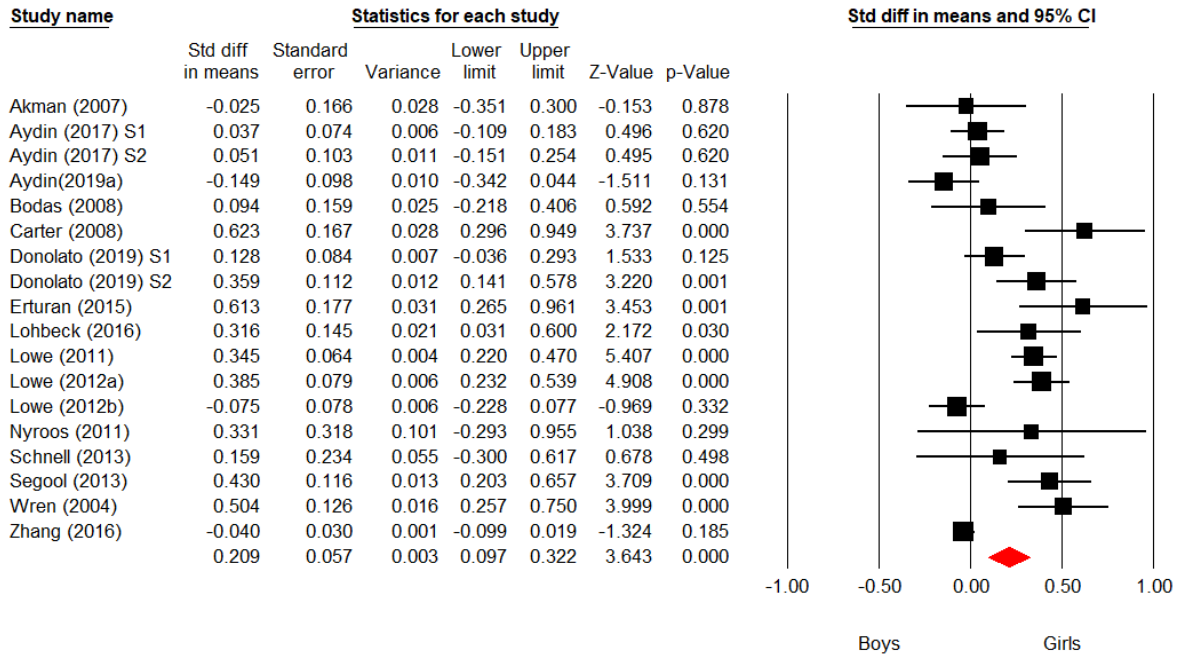


Figure 3

Meta-Analysis of gender and test anxiety scores (a) Forrest plot of standardised difference in means and 95% confidence intervals, and (b) Funnel plot of standard error by standardised difference in means.

(a)



(b)

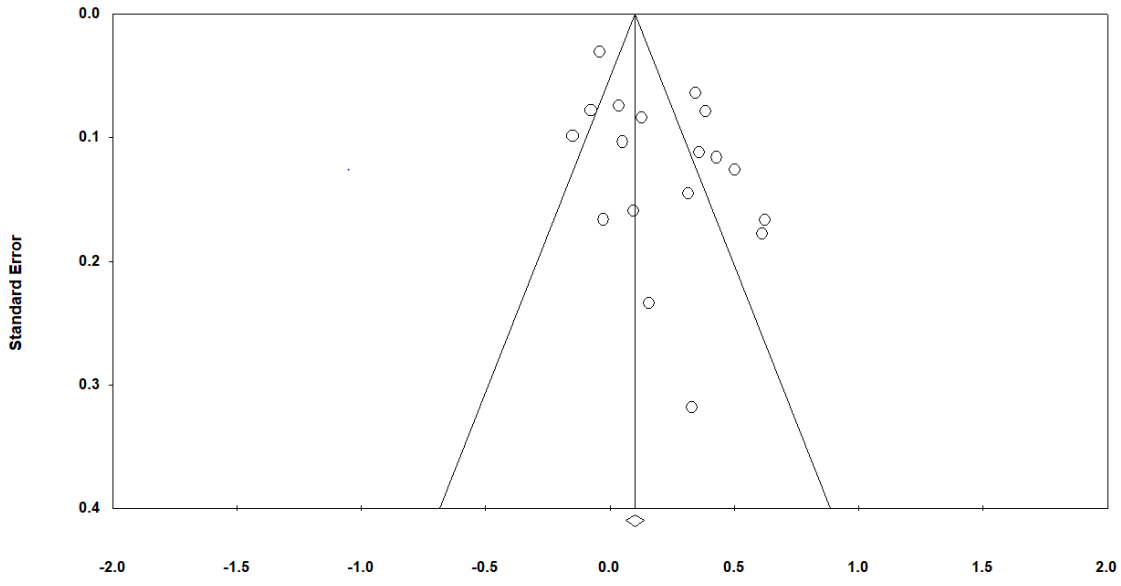


Figure 4

Graphical representation of moderation analysis for association gender differences in test anxiety scores.

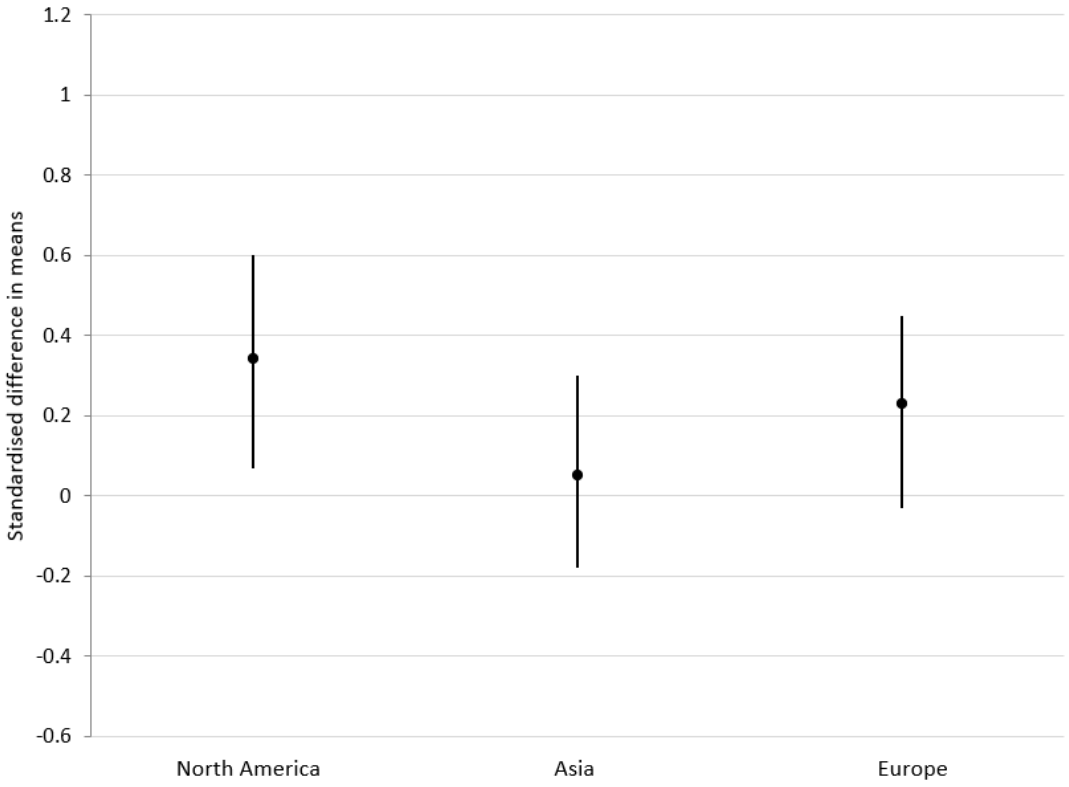
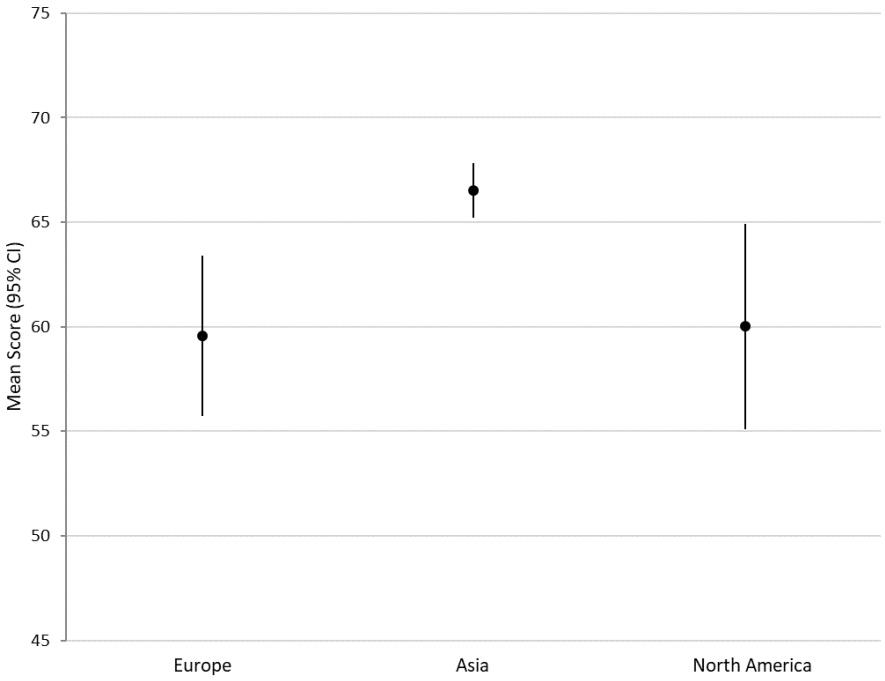


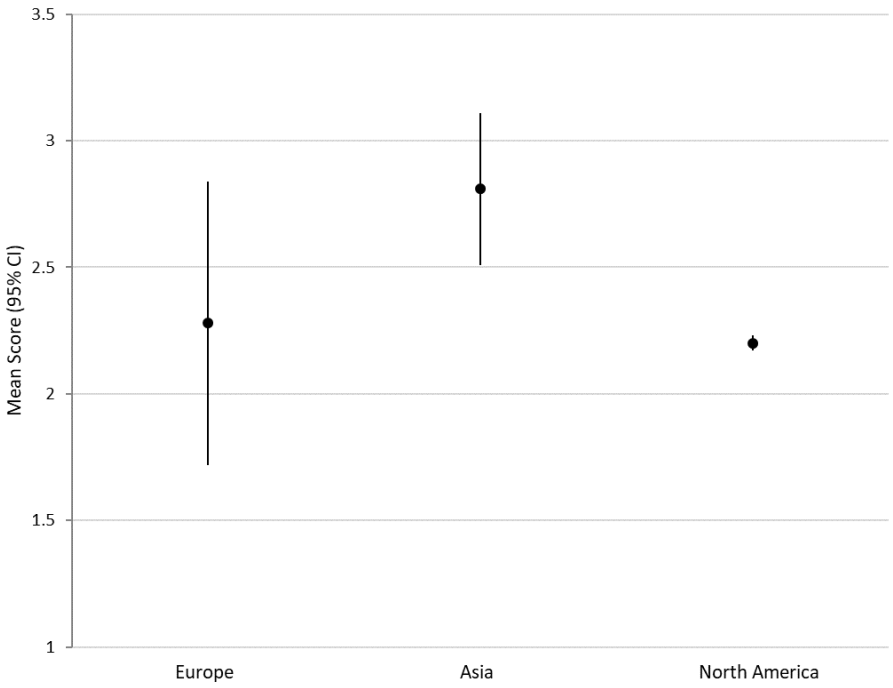
Figure 5

Graphical representation of moderation analysis of mean test anxiety scores across world regions for (a) the CTAS and (b) the MSLQ

(a)



(b)


3.5
3
2.5
2
1.5
1
Mean Score (95% CI)
Europe
Asia
North America

Test anxiety in Primary School Children: A 20 year Systematic Review and Meta-Analysis

- Supplemental Materials -

This document contains materials designed to supplement the main text. The materials include the following:

List the Tables that follow:

Table S1
Search strategy

Databases	Search Results
Scopus: Title, abstract, keywords	1699
WOS: Topic	2298
ERIC (proquest) NOFI	687
A+Education: Simple search with test anxiety with all fields	47
PsycARTICLES (Select a field-default settings)	52
PsycINFO (Select a field optional setting)	1702
MEDLINE: AB Abstract (limit to abstract available)	465
Total with duplicates	6950
Duplicates	2547
Total	4403
Scopus : Title, abstract, keywords	1699
WOS: Topic	2298

Note. The search was limited to from year 2000, full text and articles published in the English language.

Table S2
Risk of Bias

Authors	**Questions (Scored as 0 'no' or 1 'yes')**																				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Total**
Akman et al., (2007)	1	1	0	1	1	0	0	1	1	1	0	0	0	0	1	1	0	0	0	0	9
Arens et al., (2017)	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0	0	14
Aydin (2012)	1	1	0	1	0	0	0	1	1	1	0	1	0	0	1	1	1	1	0	0	11
Aydin (2013)	1	1	0	1	0	0	0	1	1	1	0	1	0	0	1	1	1	0	0	0	10
Aydin 2017	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	14
Aydin (2019a)	1	1	0	1	1	0	0	1	1	1	1	1	0	1	1	0	1	1	0	1	14
Aydin (2019b)	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	0	1	1	0	0	14
Bochis & Sandra (2018)	1	1	0	0	0	0	0	1	1	1	0	1	0	0	1	1	1	0	0	0	9
Bodas et al., (2008)	1	1	0	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0	1	15
Boehme et al., (2017)	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	16
Bong et al., (2012)1	1	1	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	16
Caputo (2014)145	1	1	1	1	1	0	0	1	1	1	0	0	0	0	1	1	1	1	1	1	14
Carey et al., (2017)144	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	17
Carsley et al., (2015)	1	1	0	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1	0	1	14
Carter et al., (2008)	1	1	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	16
Chen (2014)	1	1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	1	0	11
Datu & Fong (2018)	1	1	1	0	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	14
Donolato et al., (2019)	1	1	1	1	0	0	0	1	1	1	1	1	0	0	1	1	1	1	0	1	14
Erturan & Jansen (2015)	1	1	0	0	0	0	0	0	1	1	0	1	0	0	1	1	1	0	0	0	8

Ferrando (2015)	1	1	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	0	10
Gáspár (2001)	1	1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0	9
Goetz et al., (2007)	1	1	1	0	1	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	14
Hocaoglu & Çerkez, (2017)	1	1	1	0	1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	0	11
Karing et al., (2015)	1	1	1	0	1	0	0	1	0	1	0	1	0	0	1	1	1	1	0	0	11
Kaya (2004)	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	0	0	0	14
Khng (2017)	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	0	1	1	15
Kiili & Ketamo (2018)	1	1	0	0	0	0	0	1	1	1	1	1	0	0	1	0	1	1	1	0	11
Korhonen et al., (2018)	1	1	1	0	1	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	15
Larson et al., (2010)	1	1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	1	11
Li & Chung, (2009)	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	16
Lichtenfeld et al., (2012)	1	1	0	1	0	1	0	1	1	1	1	1	0	0	1	1	1	1	0	1	14
Ling et al., (2017)	1	1	1	1	0	0	0	1	1	1	0	0	0	0	1	1	1	0	1	0	11
Linnenbrink, (2005)	1	1	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	0	0	9
Lohbeck et al., (2016)	1	1	0	0	1	0	0	1	1	0	1	0	0	0	1	0	1	1	1	1	11
Lowe (2019)	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	0	1	15
Lowe & Ang (2012)	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	0	0	14
Lowe et al., (2011a)	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	0	0	14
Lowe et al., (2011b)	1	1	1	1	0	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	15
Mammarella et al., (2018)	1	1	1	0	1	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	15
Mavilidi et al., (2020)	1	1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	12
Mavilidi et al., (2014)	1	1	0	0	0	0	1	1	1	1	1	1	0	0	1	1	1	0	0	1	12
Metallidou, & Vlachou (2007)	1	1	0	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	12

Miller et al., (2007)	1	1	1	0	0	1	0	1	1	0	0	0	0	0	1	0	1	1	0	1	10
Miller et al., (2005)	1	1	1	0	0	1	0	1	1	0	0	0	0	0	1	0	1	1	0	1	10
Mulvenon et al., (2005)	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	1	7
Narendra et al., (2019)	1	1	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	1	1	1	9
Ng & Lee (2010)	1	1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0	9
Ng & Lee, (2015)	1	1	0	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	16
Ng & Lee, (2016)	1	1	1	1	0	0	0	1	1	1	1	1	0	0	1	1	1	1	0	1	14
Niusha et al., (2012)	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0	6
Nyroos et al., (2015a)	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	19
Nyroos et al. (2015b)	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	0	1	1	1	1	13
Nyroo & Wiklund-Hörnqvist (2011)	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1	0	1	12
Owens et al., (2012)	1	1	0	0	0	0	0	1	1	1	0	1	0	0	1	1	1	1	1	1	12
Peleg-Popko, (2002)	1	1	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	14
Putwain & Best (2011)	1	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	14
Putwain & Best (2012)	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	15
Putwain et al., (2013)	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	15
Raju & Asfaw, (2009)	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	0	0	0	13
Rouxel (2001)	1	1	1	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0	10
Schnell et al., (2013)	1	1	0	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1	0	1	12
Segool et al., (2013)	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	1	13
Selkirk et al., (2011)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
Shih (2005)	1	1	1	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	0	1	11
Shoahosseini & Baghaei, (2020)	1	1	0	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	13

Shores et al., (2010)	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0	1	1	0	1	11
Thompson et al., (2016)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
Veenman et al., (2000)	1	1	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	0	0	0	9
Villares, et al. (2014)	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	0	1	1	15
Vogelaar et al., (2017)	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	16
Wood et al., (2016)	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	16
Wren & Benson, (2004)	1	1	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	0	1	14
Yeo et al., (2016)	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	17
Zhang et al., (2016)	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0	1	16
Zhou & Urhahne (2013)	1	1	0	0	0	0	0	1	0	1	1	1	0	0	1	1	1	1	0	1	11
Zhu & Urhahne (2014)	1	1	1	0	1	0	0	1	0	1	1	0	0	0	1	1	1	1	0	0	11

Note. Computed using: Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open, 6*(12), e011458. doi:10.1136/bmjopen-2016-011458

Table S3

Description and Results of Included Studies (N publications = 76)

Author/s, Year Study Design	Country, Developing status	*N*	Age (years)	Mean, standard deviation (NP = not provided)	Gender (% M)	Test Anxiety Measure/s	Quality /20
Akman et al., (2007) Observational	Turkey LDR	145	7-9*	*M* = NP , *SD* = NP	48.3%	TAI	9
Arens et al., (2017) Observational	Germany MDR	5135	11-13*	*M* = 12.88, *SD* = 0.65	47.9%	Items Helmke (1992)	14
Aydin (2012) Observational	Turkey LDR	480	9-14	*M* = 11.8, *SD* = NP	48.7%	TASC	11
Aydin (2013) Observational	Turkey LDR	477	9-14	*M* = 11.8, *SD* = NP	51.3%	TASC	10
Aydin (2017) Observational	Turkey LDR	1100	6-14*	*M* = 9.35, *SD* = NP	51.9%	CTAS	14
Aydin (2019a) Observational	Turkey LDR	414	9-10*	*M* = 9.60, *SD* = 1.35	50.5%	CTAS	14
Aydin (2019b) Observational	Turkey LDR	615	9-12*	*M* = 10.68, *SD* = NP	50.3%	CTAS	14
Bochis & Sandra (2018) Intervention	Romania MDR	25	8-9	NP	44%	CTAS	9

Bodas et al., (2008) Observational	India LDR	231	10-15	*M* = 12.52, *SD* = 1.39	50.22%	TAI , FTA	15
Boehme et al., (2017) Observational	Germany MDR	356	9-11*	*M* = 9.89, *SD* = 1.24	43.8%	TAI/H	16
Bong et al., (2012) Observational	Korea LDR	746	8-16*	*M* = 12.55, *SD* = 0.73	55.1%	MSLQ	16
Caputo (2014) Observational	Italy MDR	3372	11-14	NP	51.1%	MSLQ	14
Carey et al., (2017) Observational	UK MDR	1720	8-13*	*M* = 10.73, *SD* = 3.85	51.5%	CTAS	17
Carsley et al., (2015) Intervention	Canada MDR	52	9-12*	*M* = 10.92, *SD* = 0.82	46.2%	STAIC-S	14
Carter et al., (2008) Observational	USA MDR	152	8-13	*M* = 9.98, *SD* = 1.13	46%	TASC	16
Chen (2014) Observational	Taiwan LDR	54	6-12*	*M* = 11, *SD* = NP	-	MSLQ	11
Datu & Fong (2018) Observational	China LDR	1051	9-11*	NP	38.5%	STES adapted from AEQ	14
Donolato et al., (2019) Observational	Italy MDR	1: 123 2: 899	1: 10-11* 2a: 7–11 2b:10–16	1:*M* = 10.89, *SD* =0.39 2a:*M* = 9.2, *SD* = 0.10	1: 46.3% 2: 49.6%	1: CTAS, FTA, TAQ-C 2: TAQ-C, TASC	14

				2b:*M* = 12.4, *SD* = 0.11			
Erturan & Jansen (2015) Observational	Netherlands MDR	134	7-15	M = 11.44, SD = 0.63 (pooled)	53%	DSQ	8
Ferrando (2015) Observational	Spain MDR	875	12-14	NP	51.17%	TASC	10
Gáspár (2001) Longitudinal	Hungary MDR	478	9-14	NP	44.4%	TAI	9
Goetz et al., (2007) Experimental	Germany MDR	2059	11-15	*M* = 12.77, *SD* = 0.52	50%	State anxiety/emotions before and after exam	14
Hocaoglu & Çerkez, (2017) Observational	Northern Cyprus LDR	210	9-11*	NP	51%	WTAS	11
Karing et al., (2015) Observational	Germany MDR	572	11-13*	*M* = 12.4, *SD* = 0.40	45%	Bespoke measure	11
Kaya (2004) Observational	Turkey LDR	463	10-13	*M* = 11.04, *SD* = 0.50	50%	TAI, STAIC	14
Khng (2017)Intervention	Singapore LDR	122	10-11*	*M* = 10.67, *SD* = 0.37	52%	TAI , CTAS , STAIC	15
Kiili & Ketamo (2018) Experimental	Finland MDR	51	11-12*	*M* = 11.92, *SD* = 0.34	45.1%	CTAS	11
Korhonen et al., (2018) Observational	Sweden MDR	624	9-10	NP	48.7%	CTAS	15

Larson et al., (2010) Intervention	USA MDR	177	8-10	NP	49.2%	WTAS	11
Li & Chung, (2009) Observational	Hong Kong LDR	519	7-12	NP	52.4%	STAIC	16
Lichtenfeld et al., (2012) Observational	1: Germany MDR 2: USA MDR	1:1189 2: 163	1: 7-12 2: 7-10	NP	1: 43.2% 2: 41.7%	AEQ	14
Ling et al., (2017) Experimental	USA MDR	789	11-13*	NP	51%	AQ	11
Linnenbrink, (2005) Experimental	USA MDR	237	10-12*	NP	51.5%	MSLQ	9
Lohbeck et al., (2016) Observational	Germany MDR	192	9-12	*M* = 10.04, *SD* = 0.64	50.5%	AFS	11
Lowe (2019) Observational	USA MDR	1221	7-9	*M* = 9.41, *SD* = 1.19	41.1%	TAS-E	15
Lowe & Ang (2012) Observational	USA MDR Singapore LDR	1322	9-12	*M* = 10.31, *SD* = 0.66	51.0%	TAS-E	14
Lowe et al., (2011a) Observational	Singapore LDR	1&2: 540 3:1080	9-13	*M* = 10.37, *SD* = 0.62	1: 45.4% 2: 46.5% 3: 45.9%	TAS-E	14
Lowe et al., (2011b) Observational	USA MDR	1: 997 2: 421	1: 7-13 2: 7-12	1: *M* = 9.76, *SD* = 1.38	1:49.0% 2:51.5%	TAS-E	15

				2: *M* = 9.67, *SD* = 1.39			
Mammarella et al., (2018)	Italy	MDR	664	8-11*	*M* = 9.20, *SD* = 1.13	52.6%	TAQC, CTAS, FTA	15
Observational								
Mavilidi et al., (2020)	Australia	MDR	68	11-12	NP	NP	CATQ	12
Intervention								
Mavilidi et al., (2014)	Greece	MDR	117	11-12	*M* = 11.59, *SD* = 0.44	43.6%	CATQ	12
Intervention								
Metallidou & Vlachou, (2007)	Greece	MDR	263	10-12*	NP	49.4%	MSLQ	12
Observational								
Miller et al., (2007)	USA	MDR	71	11-12*	NP	33%	WTAS	10
Intervention								
Miller et al., (2005)	USA	MDR	36	10-11*	NP	44%	WTAS	10
Intervention								
Mulvenon et al., (2005)	USA	MDR	251	10-11*	NP	NP	Bespoke measure	7
Observational								
Narendra et al., (2019)	India	LDR	57	12-14	NP	NP	WTAS	9
Intervention								
Ng & Lee (2010)	Singapore	LDR	90	10-11	NP	NP	TAI, STAIC	9
Experimental								
Ng & Lee, (2015)	Singapore	LDR	128	11	*M* = 10.68, *SD* = 0.58	43%	TAI , STAIC	16
Experimental								
Ng & Lee, (2016)	Singapore	LDR	113	10-12	*M* = 11.25, *SD* = 0.52	47.8%	TAI , STAIC	14
Experimental								
Niusha et al., (2012)	Iran	LDR	74	11-14*	NP	0.0%	TAI	6
Intervention								

Nyroos et al., (2015) Observational	Sweden MDR	624	8-10	*M* = 9.3, *SD* = NP	49.0%	CTAS	19
Nyroos et al., (2015) Observational	China LDR Finland MDR Sweden MDR	398	8-10	NP	48%	CTAS	13
Nyroo et al., (2011) Observational	Sweden MDR	40	9-10	NP	50.0%	CTAS	12
Owens et al., (2012) Observational	UK MDR	1: 80 2: 31	1: 12-13 2: 12-13	1: *M* = 12.1, *SD* = 0.3 2: *M* = 12.4 , *SD* = 0.5	1: 40% 2: NR	CTAS	12
Peleg-Popko, (2002) Observational	Israel LDR	152	11-12	NP	53.0%	TAI, STAIC-S	14
Putwain & Best (2011) Experimental	UK MDR	40	8-11*	NP	50.0%	CTAS	14
Putwain & Best (2012) Experimental	UK MDR	39	7-11*	NP	53.8%	CTAS , STAIC-S	15
Putwain et al., (2013) Observational	UK MDR	123	10-11*	*M* = 10.07, *SD* = 0.24	46.0%	CTAS	15
Raju & Asfaw (2009) Observational	Ethiopia LDR	497	11-13*	*M* = 12.98, *SD* = 1.49	50.0%	TAS-C	13

Rouxel (2001) Observational	France MDR	476	9-11*	*M* = 10.7, *SD* = 0.83	48.3%	STAI	10
Schnell et al., (2013) Observational	Germany MDR	79	11-13*	*M* = 12.0, *SD* = 0.93	61%	PAF, TAI	12
Segool et al., (2013) Experimental	USA MDR	335	8-11*	NP	45.1%	CTAS	13
Selkirk et al., (2011) Longitudinal	USA MDR	1953	11-13*	NP	48.0%	MAQ	20
Shih (2005) Observational	Taiwan LDR	198	11-12*	*M* = 11.6 *SD* = NP	48.0%	MSLQ	11
Shoahosseini & Baghaei (2020) Observational	Iran LDR	160	8- 14	*M* = 12.88, *SD* = 1.96	45.6%	CTAS	13
Shores et al., (2010) Observational	USA MDR	761	10-12*	NP	41.9%	MSLQ, TAI	11
Thompson et al., (2016) Intervention	USA MDR	791	10-11	*M* = 10.65, *SD* = 0.35	52.0%	CTAS	20
Veenman et al., (2000), Experimental	Netherlands MDR	1: 156 2: 75	12-13	NP	NP	PMT-K	9
Villares et al., (2014) Observational	USA MDR	4342	10-11*	NP	50.5%	MSLQ	15
Vogelaar et al., (2017) Intervention	Netherlands MDR	113	7-8	*M* = 7.91, *SD* = 6.40	47.8%	CTAS	16

Wood et al., (2016) Observational	USA MDR	426 sets of twins	10-13	*M* = 11.82, *SD* = 1.08	0.0%	CTAS	16
Wren & Benson, (2004) Observational	USA MDR	491	8-12	NP	DS- 50% VS- 47%	CTAS	14
Yeo et al., (2016) Intervention	Singapore LDR	115	9-12	*M* = 10.15, *SD* = 0.50	60.9%	CTAS	17
Zhang et al., (2016) Observational	Germany MDR	4387	9-11	*M* = 10.42, *SD* = 0.50	50.2%	TAI	16
Zhou & Urhahne (2013) Observational, Longitudinal	1: Germany MDR 2: China LDR	1: 144 2: 272	1: 9-10* 2: 9-10*	1: *M* = 9.93, *SD* = 0.61 2: *M* = 9.87, *SD* = 0.65	1: 54.2% 2: 55.1%	PISA	11
Zhu & Urhahne (2014) Observational	China LDR	480	11-13*	*M* = 11.95, *SD* = 0.95	52.8%	AEQ	11

Note. LDR- Less Developed Region, MDR- More Developed Region, NP- not provided in the paper, * age range not provided, the typical age based the grade from that country was calculated. Bespoke measure- researcher devised questionnaire. Subscale acronyms as follows: AEQ = Achievement Emotions Questionnaire (Pekrun et al., 2005); AFS = Anxiety questionnaire for students (in German: Angstfragebogen fu¨r Schu¨ler [AFS]; Wieczerkowski et al. 1981); AQ = Anxiety questionnaire (Attali & Powers, 2010); CATQ- Cognitive Anxiety Test Questionnaire (Sweller & Chandler, 1994); CTAS = Child Test Anxiety Scale (Wren & Benson, 2004); DSQ = Dutch School Questionnaire (School Vragenlijst; Smits and Vorst 2013); FTA = FRIEDBEN Test Anxiety Scale for adolescence (Friedman & Bendas-Jacob, 1997); MAQ = Mathematics Anxiety Questionnaire (Wigfield & Meece, 1988); MSLQ = Motivated Strategies for Learning Questionnaire (Pintrich et al., 1991, 1993); PAF = Prüfungsangstfragebogen/ German Test Anxiety Inventory; PISA- (PISA, 2006); PMT-K = Prestatie Motivatie Test for children (Hermans, 1976); STAIC-S = State-Trait Anxiety Inventory for Children (Spielberger, 1973); State anxiety/emotions before and after exam (Goetz, 2004); STES = The Short Test Emotions Scale adapted from the Academic Emotions Questionnaire (Pekrun et al., 2005); TAI = Test anxiety inventory (Spielberger, 1980); TAQ-C = Test Anxiety Questionnaire for Children (Donolato et al., in press); TAQ-C = Test Anxiety Questionnaire for children (Donolato et al., 2019); TASC = Test Anxiety Scale for Children (Sarason et al., 1960);TAS-E = Test Anxiety Scale for Elementary Students (Lowe et al., 2011); WTAS = The Westside Test Anxiety Scale (Driscoll, 2007).

Table S4

Description and Results of Intervention Studies (N publications = 13)

Authors year	Country	Program content	Age reported	N	Time frame/ Number of sessions	Control	Mode of delivery (Teacher Researcher, Research assistant)	Program supplied	Outcome measure/s	Key results
Bochis & Sandra (2018)	Romania	Interactive story to develop personal skills of positive thinking and relaxation. Additionally, after school training of homework and examination preparation.	8-9	25	6 weeks x 3 times a week	Non-randomised intervention with no control group	Researcher	No	CTAS (Wren and Benson, 2004)	Overall test anxiety decreased after the intervention. A significant change in the test anxiety subscale of thoughts after the intervention. No change in automatic reactions and distracting behaviour post-intervention.
Carsley et al., (2015)	Canada	Either one of two colouring conditions were given before a	10-11 *M* = 10.92 (*SD* = .82)	52	1 session (15 minutes	Randomised intervention with active	Researcher and research assistants	Mandala example supplied	STAIC-S; (Spielberger, 1973)	Both colouring conditions showed a significant decrease in

		spelling test in the classroom a) structured mindfulness based mandala colouring (intervention) b) free/unstructured colouring (control)			of colouring)	control (free colouring)				anxiety pre and post colouring. A significant interaction between gender and condition was found in both genders for the mandala group. Whereas in the free colouring condition only males decreased in anxiety.
Khng (2017)	Singapore	The intervention group learned a structured breathing technique that students did before the tests and during a 10 minute break. Research assistants helped the intervention group by sitting and guiding the students, one researcher between two students.	10-11 (*M* = 10.67, *SD* = .37)	122	1session (no time frame given)	Randomised intervention with control group	Researcher with trained research assistants	No	TAI (Spielberger, 1980) CTAS (Wren and Benson, 2004) STAIC-S (Spielberger, 1973) CCAQ (Zatz & Chassin, 1985), A computerised flanker-	Maths test results and state anxiety improved significantly for both the control and intervention groups. The intervention groups results showed a higher decrease in anxiety and larger increase in the maths score. Only boys showed a significant decrease in anxiety from the intervention group. With girls showing both a significant reduction in anxiety in

									like, distractor interference task (Forster & Lavie, 2008). WIAT–III; (Weschler, 2009)	both intervention and control group. Students with high scores on autonomic-reactions (CTAS) improved significantly better in the intervention compared to the control. No significant results were found for inhibitory-control-of-attention. Participants who did not engage in the breathing task were excluded from analysis.
Larson et al., (2010)	US	Relaxation technique involving deep breathing exercises (elevator breathing) and progressive muscle relaxation (guided relaxation).	8-10	177	5 weeks x 2 days a week (15 minutes)	Non-randomised intervention with control group	Researcher	Yes	WTAS (Driscoll, 2007)	There was a significant effect for differences in pre and post test anxiety scores for the experimental group compared to the control with test anxiety decreasing significantly in the intervention group.

Mavilidi et al., (2020)	Australia	One week before the intervention children were tested and categorised into high or low anxious groups. Randomised groups were formed with an equal mix of low and high anxious students for the condition/experimental group and a control. The experimental group participated in a 10 minute physical activity session before the maths exam. This involved activities such as push-ups, star jumps, burpees and running on the spot. The control group played a vocabulary game of “hangman” in the 10 minutes.	11-12	68	1 session (10 minutes)	Randomised intervention with control group	Researcher	Details provided in paper to the physical activities used.	CATQ (Sweller & Chandle, 1994) start of the test, during the test and after the test. Maths test	The intervention group did not have a lower test anxiety score or higher maths test performance compared to the control group. There was a significant interaction between anxiety level and time with high anxious students starting the test more anxious with their levels changing by the end of the test.

Mavilidi et al., (2014)	Greece	Strategy of looking forward/pre-reading in a problem solving maths test.	11-12 (*M* = 11.59, *SD* =0.44)	117	1 session	Randomised intervention with control group	Researcher	No	CATQ (Cassady and Johnson, 2002) Maths test Task Complexity (Sweller et al., 2011) Mental Effort Scale (Pass, 1992)	The main effect of anxiety level and test strategy, and the interaction of test strategy and anxiety level was not significant. Anxiety was not reduced based on the use of looking ahead and non-looking ahead conditions. Anxiety from the beginning to the end of the test did significantly decrease for both conditions. Student who used the looking ahead strategy did significantly better in the maths test with a moderate-large effect size.
Miller et al., (2007)	US	Children were pre-tested for anxiety, children with high anxiety were chosen for the program. The	11-12*	71	2 week period with 3 sessions (1 live and 2 pre-	Randomised intervention with control group	Researcher	A brief outline of the pre-recorded content	WTAS (Driscoll, 2007)	There was a significant effect for the treatment group with the invention group reporting lower scores in test anxiety.

		students were randomly assigned to the control or intervention group. The pre-recorded program included guided relaxation activities, also imagined scenarios based on accelerated desensitisation and adaptive attitudes. For example, students imagined positive emotions to schoolwork in difference scenarios and exposure to testing.			recorded) The recording was given to the students to review at home. Same sex groups made.			of the guided sessions was provided in the paper.	Academic achievement test	There were no differences in academic achievement scores between the treatment groups.
Miller et al., (2005)	US	Children were tested for anxiety and children with high anxiety were chosen for the program. The pre-recorded program included guided relaxation activities, also imagined scenarios based on accelerated desensitisation and adaptive attitudes.	10-11*	36	5 pre-recorded (31 minutes) sessions over half the school year. Same sex groups made.	Randomised intervention with control group	School counsellor	A brief outline of the pre-recorded content of the guided sessions was provided in the paper.	Academic achievement test	Test anxiety measures were taken to place students in the program. There was a significant effect for the treatment group with the invention group reporting higher scores on an achievement The intervention group demonstrated significantly better

		For example, students imagined positive emotions to schoolwork in difference scenarios and exposure to testing.								scores in reading and social studies. There was no significant differences between the intervention and control group in mathematics and science scores.
Narendra et al., (2019)	India	Yoga exercises which included breathing exercises, physical poses, and guided relaxation.	12-14	57	6 weeks x 5 days a week (40 minutes sessions)	Non-randomised intervention with active control group Active control group (education pamphlet on reducing anxiety)	Researcher	A list of the yoga positions and the timing for each one is provided in the paper.	TAI	The control group and the intervention group both decreased significantly in anxiety from pre-test to post-test scores.
Niusha et al., (2012)	Iran	Assertiveness training using role play for female students.	11-14*	74	8 x 50 minutes (no time frame given)	Non-randomised intervention with control group Schools were randomly chosen	Researcher	No	TAI measured after course, 2 and 4 months later	There was a significant main effect for treatment. The experimental group reporting lower TA scores than the control group. The test anxiety score remained stable after 2 and 4 months after the intervention.

Thompson et al., (2016)	US	Moderate-to-vigorous physical education lesson before test taking.	10-11 (*M* = 10.66 *SD* = 0.35)	791	1 session (20-40 minutes before a test)	Randomised intervention with control group	PE teachers	Activities completed in the PE lessons are published in the paper.	CTAS Teacher survey of test taking behaviour Academic assessment	No statistically significant differences in test anxiety total scores or subscales between the experimental and control groups. No differences between mathematics and reading test scores between the intervention and control groups. Teacher survey results showed better behaviour and better overall test taking behaviour in the maths test in the intervention group. No differences were found in the reading tests between control and intervention groups. There was a difference in sex and race. African American students in the intervention group had lower CTAS off-task behaviours scores on

										mathematics and reading tests. Boys had more off-task behaviours on the reading test in the intervention group.
Vogelaar et al., (2017)	Netherlands	The dynamic training session used geometric analogy problems to build problem solving strategies. The program had graduated metacognitive and cognitive prompt techniques administer hierarchically to help solve geometric problems.	7-8 (*M* = 7.91, *SD* = 6.40)	113	1 session	Randomised intervention with active control group Control group did a dot-to-dot activity Randomised bases on pairs of children with equal Raven scores (blocking).	Researcher	A general procedure is published in the article.	CTAS;	Children with higher test anxiety of improved with number of correct analogies more across test sessions than children with low test anxiety. Gifted and average ability children were not differentiated by test anxiety.
Yeo et al., (2016)	Singapore	The program was based on CBT interventions for anxiety disorders. Modules included psycho-education, relaxation training,	9-12 (*M* = 10.15, *SD* = 0.50)	115	4 weeks x 4 days (30 mins)	Non-randomised intervention with control group	Researcher	Detailed activities on the lessons are publishe	CTAS	No significant change in test anxiety for CBT participants was found between baseline and post-treatment.

	self-instruction, exposure, and skills training components. Homework relaxation exercises were assigned, students given an audio CD with a script.		d in the article.	The mean change in anxiety scores from pre-intervention to follow-up was significantly greater for the CBT group than the control group. CBT group reported significantly lower test anxiety at 2 months follow-up compared to baseline and posttreatment. Significant differences were found at post-treatment and at follow-up for only the High Severity group.

Table S5

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Included
TITLE			
Title	1	Identify the report as a systematic review.	☑
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	☑
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	☑
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	☑
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	☑
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	☑
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	☑
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	☑
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	☑
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	☑
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	☑
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	☑
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	☑
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	☑

Section and Topic	Item #	Checklist item	Included
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	☑
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	☑
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	☑
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	☑
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	☑
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	☑
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	☑
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	☑
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	☑
Study characteristics	17	Cite each included study and present its characteristics.	☑
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	☑
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	☑
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	☑
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	☑
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	☑
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	☑
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	☑
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	☑

Section and Topic	Item #	Checklist item	Included
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	☑
	23b	Discuss any limitations of the evidence included in the review.	☑
	23c	Discuss any limitations of the review processes used.	☑
	23d	Discuss implications of the results for practice, policy, and future research.	☑
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	☑
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	☑
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	☑
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	☑
Competing interests	26	Declare any competing interests of review authors.	☑
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	☑