Journal of Psychopathology and Behavioral Assessment (2026) 48:23
https://doi.org/10.1007/s10862-026-10285-5

# Factor Structure, Measurement Invariance, and Psychometric Properties of the Reactions to Tests for Adolescents in a Portuguese Sample

**Cláudia P. Pires**[1,2,3] · **David W. Putwain**[4] · **Stefan G. Hofmann**[5] · **Ana Rita Vicente**[6] · **Maria do Céu Salvador**[3]

Received: 30 September 2024 / Accepted: 6 April 2026 / Published online: 1 May 2026


## Abstract

Adolescent test anxiety is concerning, yet few multidimensional tools with established cross-cultural validation and international use exist to assess it. This study validated Sarason's Reactions to Tests in a large Portuguese adolescent sample (RTT-A) and critically examined the empirical and theoretical relevance of its most debated dimension, Test-Irrelevant Thinking, whose status as a core cognitive component of test anxiety versus a more general attentional disengagement process remains contested. A total of 1395 Portuguese adolescents (57.9% female, $M_{age}$ = 15.83) completed the RTT-A, with 299 also completing additional anxiety measures, and 120 completing a retest after five weeks. Two competing models – a four-factor solution including Test-Irrelevant Thinking and a three-factor model excluding it – were tested via confirmatory factor analysis. Both models showed good fit, strong internal consistency, test–retest reliability, and measurement invariance across gender. Convergent and divergent validity were supported by moderate to strong correlations with proximal constructs (physical symptoms, social anxiety, and performance anxiety in formal social situations), and weaker associations with more distal constructs (separation anxiety and harm avoidance). Notably, Test-Irrelevant Thinking consistently displayed lower factor loadings, weaker correlations with related constructs, and no incremental validity in predicting social or performance anxiety, questioning its role as a core component of test anxiety. These findings suggest it may reflect an attentional disengagement process rather than an anxiety-specific mechanism. Nonetheless, both RTT-A versions demonstrated sound psychometric properties, and utility for cross-cultural research and practice. The three-factor version may offer a more parsimonious assessment of test anxiety's cognitive and affective-physiological components, while the four-factor model may be informative in studies exploring attentional or avoidance-related processes.

**Keywords** Test anxiety · Adolescents · Cross-cultural validation · Confirmatory factor analysis · Test-Irrelevant Thinking · Measurement invariance

✉ Cláudia P. Pires
claudia.pires@ulusofona.pt

1 HEI-Lab – Digital Human-Environment and Interaction Labs, Universidade Lusófona, Lisbon, Portugal

2 Instituto Superior Miguel Torga, Coimbra, Portugal

3 Center for Research in Neuropsychology and Cognitive Behavioral Intervention (CINEICC), Faculty of Psychology and Education Sciences, University of Coimbra, Coimbra, Portugal

4 School of Education, Liverpool John Moores University, Liverpool, UK

5 Philipps-University Marburg, Marburg, Germany

6 Inclusion Resources Center, Association for Education of Maladjusted Children, Torres Vedras, Portugal

Test anxiety refers to the cognitive, behavioural, or emotional responses triggered by a deep fear of poor performance in evaluative situations such as tests or exams (Zeidner, 1998). Literature reports varying prevalence rates of test anxiety, often dependent on cutoffs used in self-report scales. Studies indicate that the percentage of students reporting high levels of test anxiety ranges from 11 to 39%, spanning from elementary school through university (e.g., Gerwing et al., 2015; Putwain & Daly, 2014; Segool et al., 2013).

Because test anxiety tends to intensify during school transitions and high-stakes evaluations (Banks & Smyth, 2015), it is particularly relevant to examine this phenomenon during adolescence – a developmental period marked by significant neurobiological and psychosocial changes,

typically ranging from ages 11 to 19, though some definitions extend to 24 or 25 depending on brain maturation and social role transitions (Johnson et al., 2009; Salmela-Aro, 2011; Sawyer et al., 2018). While the recognition of this life stage is universal, its conceptualization and lived experience vary across cultural contexts. In Western societies, including Portugal, adolescence is often framed as a phase of increasing autonomy and individual achievement, whereas in collectivistic cultures, it may be characterized by stronger interdependence with family and community (Gibbons & Poelker, 2019). Despite these differences, adolescents across the globe face stressors across academic, social, and familial domains, which heighten their vulnerability to internalizing difficulties such as anxiety (Grant et al., 2014). Although academic stress is often more pronounced in East Asian contexts due to strong parental and societal expectations (Ang et al., 2009), Western performance-oriented educational systems, such as the Portuguese, also exert considerable pressure through high-stakes assessment practices (Banks & Smyth, 2015; Levesque, 2018). In Portugal, for instance, formal written evaluations in the 9th to 12th grades (normally aged 14–18) often play a decisive role in shaping adolescents' academic and professional futures. These pressures, though culturally variable in origin, consistently place adolescents at elevated risk for test anxiety during a developmentally sensitive period marked by intense cognitive, social, and emotional demands (Bound et al., 2009; Conner et al., 2009; Feld & Shusterman, 2015).

Adolescent students with high test anxiety are more likely to have lower academic performance (Putwain et al., 2016a, 2016b), experience academic burnout (Fisher et al., 2023) and elevated levels of distress, including depression and anxiety (Wuthrich et al., 2021). They also report lower life satisfaction, reduced well-being (Steinmayr et al., 2016), and physical health issues such as sleep disturbances during adolescence (Chamberlain et al., 2011). In Portugal, for example, university students frequently seek help from counselling centers for test and exam anxiety, which often started during secondary school (Melo, 2006). The severity and impact of test anxiety underscore the need for validated assessment tools to enable effective, targeted interventions.

In recent decades, conceptualizations of test anxiety evolved from early unidimensional models (Mandler & Sarason, 1952; Zeidner, 1998) to more nuanced approaches that distinguish cognitive and affective-physiological components (Liebert & Morris, 1967; Sarason et al., 1960, 1964; Spielberger et al., 1978). A key theoretical development during this period was the cognitive interference model proposed by Wine (1971), which emphasized the disruptive role of intrusive cognitions on performance. According to this model, anxious individuals split their attention between task-relevant activities and intrusive cognitions, such as self-criticism or worries about failure, which interfere with test performance (Wine, 1971). This perspective laid the foundation for later multidimensional conceptualizations of test anxiety. Building on these insights, Sarason (1984) developed one of the most influential multidimensional models of test anxiety: the Reactions to Tests (RTT) framework. This 40-item self-report instrument comprises four dimensions – Tension (awareness and interpretation of physiological changes and arousal), Bodily Symptoms (physiological symptoms of anxiety), Worry (distracting worrying thoughts related to test performance), and Test-Irrelevant Thinking (frequency and intensity of thoughts unrelated to the test situation). The first three are widely accepted as core components of test anxiety, mapping onto its affective-physiological (Tension and Bodily Symptoms) and cognitive (Worry) dimensions (e.g., Liebert & Morris, 1967; Putwain et al., 2021a, 2021b; Spielberger & Vagg, 1995; Spielberger et al., 1978). In contrast, the validity of the Test-Irrelevant Thinking dimension has been increasingly debated (Putwain et al., 2021a, 2021b; Rost & Schermer, 1992). Originally proposed by Sarason (1984) to capture thoughts unrelated to the test situation (e.g., *Irrelevant bits of information pop into my head during a test*, *My mind wanders during tests*, *I daydream during tests*), this dimension reflects cognitive disengagement that lacks the evaluative and emotional tone of Worry. Whereas worry represents anxiety-related cognitive interference, test-irrelevant thinking seems to capture a more diffuse attentional disengagement, showing weaker associations with physiological arousal and other RTT dimensions (e.g., Baptista et al., 1989; Benson & Bandalos, 1992; Sarason, 1984; Pereira et al., 2021; Vicente & Salvador, 2011).

Some studies suggest that this attentional disengagement may overlap with the phenomenon of unintentional mind-wandering – a spontaneous, non-deliberate shift of attention away from task-relevant content (Smallwood & Schooler, 2014) – that reflects momentary lapses in executive control and is associated with poorer task performance (Seli et al., 2016). Unlike anxiety, typically characterized by heightened threat perception and amygdala hyperreactivity (Shin & Liberzon, 2010), this process reflects a transient failure of attentional control rather than an emotional or cognitive response to threat. This framework offers a plausible explanation for why the test-irrelevant thinking dimension is weakly related to the affective and cognitive components of test anxiety, suggesting that it may represent an attentional process that co-occurs with, rather than defines, anxiety. Empirical evidence supports this interpretation, showing that unintentional mind-wandering is associated with both anxiety and attentional dysfunctions (Figueiredo et al., 2020; Seli et al., 2019). Accordingly, Test-Irrelevant Thinking may index a broader attentional vulnerability that

amplifies the experience of test anxiety, rather than constituting a core component of it.

Furthermore, the test-irrelevant thinking dimension appears conceptually narrower than the other RTT factors, as it focuses almost exclusively on cognitive disengagement during test-taking, overlooking other relevant phases such as test preparation – a well-established source of test anxiety (Chou, 2018; Putwain et al., 2009; Salvador, 2009). In contrast, the other RTT dimensions explicitly capture pre-test experiences (e.g., *I feel distressed and uneasy before tests* – Tension; *Before tests, I feel troubled about what is going to happen* – Worry; *I get a headache before a test* – Bodily Symptoms). This specificity may limit the generalizability of the test-irrelevant thinking factor, as such disengagement experiences might characterize only a subset of test-anxious students.

These attentional processes may indeed not affect all individuals equally, as developmental and neurocognitive factors – including gender-related patterns of attentional control and emotional regulation – may shape how test anxiety is experienced and expressed. In fact, across developmental stages, female students report consistently higher levels of test anxiety than their male counterparts (e.g., Hembree, 1988; Lowe, 2019; McDonald, 2001; Núñez-Peña et al., 2016; Putwain & Daly, 2014; von der Embse et al., 2018), including in Portuguese samples (Baptista et al., 1989; Cunha & Paiva, 2012; Salvador, 2009). Although these differences are often attributed to sociocultural factors such as internalizing tendencies and academic pressure (Ben-Zur & Zeidner, 2012; di Maria & di Nuovo, 1990; Egloff & Schmukle, 2004; Zeidner, 2014), they may also reflect neurocognitive mechanisms. For instance, inattentive symptoms, more prevalent among females with Attention-Deficit/Hyperactivity Disorder (ADHD) (Rucklidge, 2010), have been linked to impaired working memory and sustained attention (D'Agati et al., 2019; Salari et al., 2023). When co-occurring with anxiety, these symptoms may heighten cognitive interference and contribute to the gender gap in self-reported test anxiety (Arnold et al., 2015; Owens et al., 2014). This integration of psychosocial and attentional factors strengthens the rationale for examining both affective and cognitive dimensions of test anxiety, including the debated role of Test-Irrelevant Thinking.

Beyond gender differences, research has also explored how test anxiety relates to broader domains of psychological functioning (see von der Embse et al., 2018, for a review). Although extensively studied within educational psychology, test anxiety is not formally recognized as a distinct disorder in current diagnostic systems such as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR; American Psychiatric Association [APA], 2022) or the International Classification of Diseases for Mortality and Morbidity Statistics (ICD-11; World Health Organization [WHO], 2019). Nonetheless, its core features – persistent fear of negative evaluation, excessive worry, and physiological hyperarousal – closely parallel those observed in anxiety disorders, particularly social anxiety.

Notably, several authors have conceptualized test anxiety as a context-specific expression of broader social-evaluative fears (Bögels et al., 2010), given that both share heightened self-focused attention and sensitivity to perceived judgment. Empirical findings consistently support this overlap: higher levels of social anxiety predict higher test anxiety (e.g., Kavanagh et al., 2017; Salvador, 2009). This conceptual link was reflected in earlier diagnostic frameworks, as the DSM-IV (APA, 1994) explicitly listed test-taking among the performance situations feared in social phobia (now social anxiety disorder), together with public speaking and eating in front of others. At the same time, test anxiety shows consistent associations with somatic symptoms such as increased heart rate, nausea, dizziness, and restlessness – responses typical of anxiety activation (Roos et al., 2021, 2022; Sarason, 1984; Zeidner, 1998) – yet it also displays distinctive correlational patterns compared to other anxiety forms. For instance, Laurin-Barantke et al. (2016) found that social anxiety predicted test anxiety primarily in oral exams, but not in written ones, suggesting partial but not complete overlap. Similarly, Knappe et al. (2011) reported that many adolescents who met diagnostic criteria for social anxiety disorder did not fear taking tests, and vice versa.

Moreover, test anxiety has shown weaker associations with anxiety subtypes less directly tied to evaluative contexts, such as those reflecting general anxious coping tendencies (e.g., hypervigilance, avoidance of distress, or repetitive safety-checking behaviors) (Hasani et al., 2025; Putwain et al., 2021a, 2021b). In line with this distinction, factor-analytic studies have demonstrated that test anxiety is empirically separable from both generalized anxiety and panic disorder symptoms (Putwain et al., 2021a, 2021b). Similarly, it tends to show weaker associations with anxiety profiles more rooted in interpersonal safety and attachment, such as separation anxiety (Hasani et al., 2025) or social connectedness (Kavanagh et al., 2017). These differential associations highlight the conceptual specificity of test anxiety and underscore the importance of examining its relationship to other anxiety domains.

Until the turn of the twenty-first century, most test anxiety measures were developed for adult populations, particularly university students. However, more recent efforts have produced adolescent-focused instruments that reflect developmental and contextual factors. These include the 23-item FRIEDBEN Test Anxiety Scale (FTAS; Friedman & Bendas-Jacob, 1997), its 12-item brief version (von der Embse et al., 2013), and the biopsychosocial model-based

instruments from Lowe and colleagues: the Test Anxiety Inventory for Children and Adolescents (TAICA; Lowe et al., 2008) and the Test Anxiety Scale for Elementary Students (TAS-E; Lowe et al., 2011). These measures incorporate behavioural, emotional, cognitive, and social dimensions, thus supporting the view that test anxiety is influenced by multiple interacting factors, including social fears and expectations (e.g., Bögels et al., 2010; Kavanagh et al., 2017; Laurin-Barantke et al., 2016). Nonetheless, and as mentioned, it has been argued that social and behavioural components – as well as cognitive distractibility, such as test-irrelevant thinking – may act as antecedents or correlates of test anxiety, rather than constituting part of its core definition, which is more commonly defined in terms of cognitive and affective-physiological responses to perceived evaluative threat (Putwain et al., 2021a, 2021b).

In evaluating test anxiety within the Portuguese context, the availability of robust assessment tools remains somewhat limited. Most available instruments have been developed for university students. For instance, the short version of the widely used Test Anxiety Inventory-5 (TAI-5; Taylor & Deane, 2002) was adapted for Portuguese university students by Ponciano et al. (2005). Similarly, the RTT was validated for Portuguese university students, replicating the four-factor structure, retaining 34 items, and demonstrating good psychometric properties (Baptista et al., 1989). A short form of the RTT (Benson & Bandalos, 1992) was more recently validated in medical students, retaining the four factors and showing gender invariance (Pereira et al., 2021). For adolescents, Silva et al. (2017) adapted the Cognitive Test Anxiety Scale-Revised (Cassady & Finch, 2015), and Rosário and Soares (2004) developed the Test Anxiety Questionnaire specifically for Portuguese adolescents, proposing a multidimensional trait-based framework emphasizing competitive thoughts and tension. Additionally, Vicente and Salvador (2011) conducted an exploratory factor analysis of the RTT in a sample of Portuguese adolescents, successfully replicating the original four-factor structure and item content observed in adult populations.

Despite these efforts, limitations remain. The study by Vicente and Salvador (2011), while promising, relied exclusively on exploratory methods; it did not test the model using confirmatory techniques, nor did it evaluate alternative structures – such as a three-factor version excluding the debated Test-Irrelevant Thinking dimension. Moreover, that study did not examine gender differences, and important psychometric properties such as factorial invariance, convergent and divergent validity, and test–retest reliability were not assessed, all of which are necessary to confirm measure adequacy (Thompson, 2004). The Cognitive Test Anxiety Scale-Revised (Cassady & Finch, 2015; Silva et al., 2017), in turn, does not address affective-physiological components of test anxiety, and the Test Anxiety Questionnaire by Rosário and Soares (2004), although multidimensional, has seen limited use outside Portugal and has not undergone cross-cultural validation or achieved broader visibility within the test-anxiety literature, limiting its utility in comparative research. As Zhao et al. (2024) highlight, measurement instruments tested in only one cultural context often fail to demonstrate conceptual or psychometric equivalence in others, underscoring the importance of cross-culturally validated tools to ensure reliable comparisons and valid inferences.

In line with this, the RTT emerges as a particularly relevant instrument to validate for Portuguese adolescents. It is widely used and internationally recognized within test-anxiety research, grounded in a multidimensional framework, and has already shown promising results in prior research – including Portuguese research (Baptista et al., 1989; Benson & Bandalos, 1992; Pereira et al., 2021; Vicente & Salvador, 2011). Additionally, given ongoing debates concerning the conceptual distinctiveness and construct validity of the Test-Irrelevant Thinking dimension, it is particularly important to compare the original four-factor model with an alternative three-factor structure that omits this component. Therefore, validating the RTT-A through confirmatory factor analysis (CFA) in a Portuguese adolescent sample may contribute to establishing a psychometrically sound instrument, facilitate cross-study comparability, and offer additional insights into the empirical structure and dimensionality of the test anxiety construct.

## Purpose of the Present Study

The key aims of this study were to validate the RTT for Portuguese adolescents (RTT-A) and to examine the empirical adequacy of the Test-Irrelevant Thinking dimension in this population. To do so, we tested two competing versions of the RTT-A: a four-factor model including Test-Irrelevant Thinking, and a three-factor model excluding it. These models were compared through CFA, with additional analyses of measurement invariance across gender, reliability (internal consistency and test–retest stability), and relationships with other anxiety-related variables. The incremental validity of the Test-Irrelevant Thinking factor was also examined by testing its unique contribution to the prediction of anxiety-related variables (social anxiety and anxiety specific to performance in formal social situations), beyond the variance explained by the other three RTT dimensions. Finally, we report descriptive statistics for the RTT-A total scale and subscales by gender and for the overall sample.

The formulation of our hypotheses was guided by previous empirical and theoretical work. First, we hypothesized

that the four-factor model would demonstrate good fit, based on prior validations of the RTT in both adult and adolescent populations, including Portuguese samples (e.g., Baptista et al., 1989; Pereira et al., 2021; Vicente & Salvador, 2011). Second, we hypothesized that the RTT-A would show measurement invariance across gender, drawing on evidence from the short-form RTT validation with Portuguese medical students (Pereira et al., 2021). Third, we expected good internal consistency and temporal stability, consistent with previous research reporting strong psychometric properties for both the full and abbreviated RTT versions (e.g., Baptista et al., 1989; Pereira et al., 2021). Fourth, we predicted that RTT-A scores would correlate more strongly with anxiety subtypes closely related to test anxiety (i.e., performance anxiety in formal social situations, social anxiety, and somatic symptoms) than with more distal subtypes such as separation anxiety or general anxious coping (harm avoidance), in line with prior studies emphasizing the conceptual specificity of test anxiety (e.g., Hasani et al., 2025; Kavanagh et al., 2017; Putwain et al., 2021a, 2021b). Fifth and finally, we hypothesized that girls would report higher test anxiety than boys, reflecting consistent gender differences documented across developmental stages and cultures, including Portuguese samples (e.g., Hembree, 1988; Egloff & Schmukle, 2004; von der Embse et al., 2018; Baptista et al., 1989; Salvador, 2009; Cunha & Paiva, 2012). Although prior studies have discussed alternative structures of the RTT and questioned the distinctiveness of the Test-Irrelevant Thinking dimension (e.g., Putwain et al., 2021a, 2021b; Rost & Schermer, 1992; Vicente & Salvador, 2011), empirical evidence remains too limited and inconsistent to support a clear directional hypothesis. Accordingly, we refrained from formulating specific predictions regarding the three-factor model or the incremental validity of this dimension.

## Method

### Participants

A total of 1395 Portuguese adolescents (57.9% female) aged between 14 and 18 years (grades 9 to 12), from public and private schools in central and southern Portugal, participated in this study. The mean age was 15.83 years ($SD$=1.32), and the average years of schooling was 10.35 ($SD$=1.17). No significant gender differences were found either for age, $t(1393)=0.18$, $p=.859$, $d=.01$, or years of schooling, $t(1393)=-0.76$, $p=.448$, $d=-.05$. However, students in public schools had significantly more years of schooling than those in private schools, $t(1393)=2.96$, $p=.003$, $d=.22$, and were also older, $t(253.22)=4.74$, $p<.001$, Hedges' $g=.38$, with both effects being small in magnitude. Significant regional differences, although with small effects, were also found: students in southern Portugal were older, $t(1393)=-6.11$, $p<.001$, $d=-.31$, and had more years of schooling, $t(1329.44)=-6.70$, $p<.001$, Hedges' $g=-.35$, compared to those in the centre. Gender distribution also differed by region, $\chi^2(1)=11.20$, $p<.001$, $V=.09$, with a higher proportion of girls in both regions – although the effect was also small.

### Measures

All participants completed the 34-item version of the RTT-A previously used in the exploratory study by Vicente and Salvador (2011), adapted from the Portuguese version for university students (Baptista et al., 1989). This version retained the original four-factor structure – Tension, Bodily Symptoms, Worry, and Test-Irrelevant Thinking – and was designed to assess the level of adolescent anxiety in test/exam situations. Items were rated on a 4-point Likert scale from 1 (*not at all typical of me*) to 4 *(very typical of me*), higher scores indicating higher levels of test anxiety. The original version developed by I. G. Sarason (1984) demonstrated acceptable internal consistency ($\alpha=.78$ for the total scale; $\alpha$s=.68-.81 for subscales). The Portuguese university student version (Baptista et al., 1989) and the adolescent exploratory study (Vicente & Salvador, 2011) both showed excellent internal consistency ($\alpha=.93$ for the total scale; $\alpha$s=.75-.90 for subscales).

The *Social Anxiety and Avoidance Scale for Adolescents* (SAASA; original Portuguese version by Cunha et al., 2004, 2008) is a 34-item self-report instrument designed to assess the level of discomfort and avoidance in a wide range of social situations commonly associated with social fears in adolescence. It includes two subscales: discomfort/anxiety and avoidance. In the discomfort/anxiety subscale, responses range from 1 (*not at all*) to 5 (*extremely*), whereas in the avoidance subscale they range from 1 (*never*) to 5 (*almost always*). In addition to a total score, separate scores are computed for each subscale, with higher scores indicating greater social anxiety. Each subscale is further divided into six factors: (i) interaction in new social situations, (ii) interaction with the opposite sex, (iii) performance in formal social situations, (iv) assertive interaction, (v) observation by others, and (vi) eating and drinking in public. In this study, only the performance in formal social situations factor of the discomfort/anxiety subscale was used as a concurrent measure similar to test anxiety. This factor includes various classroom social evaluation scenarios, such as reading aloud in class, taking an oral test or exam, or being called to the blackboard. The SAASA demonstrated good internal consistency, with a Cronbach's alpha value of .91

for the discomfort/anxiety subscale, and .87 for the avoidance subscale in the original version. In this study, the alpha for performance in formal social situations was .81.

The *Multidimensional Anxiety Scale for Children* (MASC; March et al., 1997; Portuguese version by Salvador et al., 2017) is a 39-item measure to assess anxiety symptoms in children and adolescents. It is composed of four factors, three of them with two subfactors: (i) Physical symptoms, with the subfactors tension/restlessness and somatic/autonomic symptoms, (ii) social anxiety, subdivided in humiliation/rejection fears and performance anxiety, (iii) separation anxiety, and (iv) harm avoidance, subdivided in perfectionism and anxious coping. Participants respond to the frequency they experience the situation reflected in each item on a 4-point Likert scale from 0 (*never true about me*) to 3 (*often true about me*). The higher the score, the higher the level of anxiety. The original version showed good internal consistency for the total score and the four main factors (.74 to .90), as did the Portuguese version, with a Cronbach's alpha of .89 for the total score and between .70 and .85 for the four main factors. This version also showed good test–retest reliability and adequate convergent validity, with positive associations with other measures of anxiety and depression, and negative associations with a measure of well-being. In this study, only the physical symptoms, separation anxiety and harm avoidance factors were used as concurrent measures – the first considered conceptually more proximal to test anxiety, and the latter two less so. Cronbach's alpha values in the present sample were .92 for the total scale, .85 for physical symptoms, .74 for separation anxiety, and .72 for harm avoidance.

## Procedure

This study was approved by the host institution on November 28, 2019, and all procedures were conducted in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Permission was granted by the author of the RTT (Sarason, 1984) to validate the scale in the Portuguese adolescent population. The RTT version used in this study had previously been adapted from the Portuguese adult version considering participants' age, in the exploratory study by Vicente and Salvador (2011). A convenience sample was recruited from public and private schools in central and southern Portugal, selected due to their proximity to the research institution and logistical feasibility. Head Teachers were contacted via email to seek their collaboration. After this permission was obtained, an informed consent form was sent to the legal guardians, detailing the study information and ensuring voluntary and confidential participation. Adolescents whose guardians consented to their participation were also informed about the research and signed an informed assent form. To assess the relationship of the RTT-A with other variables, a subsample of 299 participants completed additional anxiety-related measures. A further subsample of 120 students completed the RTT-A a second time after a 5-week interval to evaluate temporal stability. This interval was selected to balance practical feasibility with methodological recommendations for minimizing recall bias while ensuring stability over time (e.g., Vaz et al., 2013).

## Data Analysis

Statistical analyses were carried out using the Statistical Package for the Social Sciences (SPSS), version 25 (IBM Corp., 2017) and Amos, version 25 (Arbuckle, 2017). Normal distribution of items was confirmed through coefficients of skewness and kurtosis ($|Sk| < 3$ and $|Ku| < 10$; Kline, 2016), and the presence of outliers through Mahalanobis distance (*D2*). Independent sample *t*-tests and chi-square tests were used to examine gender differences in age and years of schooling, as well as differences by school region (centre vs. south) and type (public vs. private). These tests were also applied to explore gender differences and school region and type differences in RTT-A's total scores and factor scores (Field, 2009). Effect sizes were interpreted according to Cohen's (1988) guidelines: *d* values of .20-.49 were considered small, .50-.79 medium, and $\geq .80$ large. For chi-square tests, effect size was assessed using Cramer's *V*, with .10-.29 indicating small, .30-.49 medium, and $\geq .50$ large effects (Cohen, 1988). Homogeneity of variances was tested with Levene's test; when this assumption was met, Cohen's *d* was used to compute effect size, otherwise Hedges' *g* was applied.

A confirmatory factor analysis (CFA) was performed to test a three- and a four-factor model, using the items obtained in the RTT's Portuguese exploratory study for adolescents (Vicente & Salvador, 2011), with Maximum Likelihood estimation. The four-factor model included all items corresponding to Tension, Worry, Bodily Symptoms, and Test-Irrelevant Thinking. The three-factor model was identical except for the exclusion of the nine items comprising the Test-Irrelevant Thinking factor, retaining only Tension, Worry, and Bodily Symptoms. The bootstrap procedure (5000 samples) was used to obtain an accurate estimation of standard errors as reflected in *p* values and 95% confidence intervals (Nevitt & Hancock, 2001). The goodness-of fit indices to verify model adjustment were chi-square, which indicates a good fit when non-significant, although it is sensible to high sample sizes; normed chi-square ($\chi^2$/df), which

indicates a good fit with values below 5; Comparative Fit Index (CFI) and Tucker and Lewis Index (TLI), which indicate a good fit when above .90; and the root mean square error of approximation (RMSEA; 90% confidence interval [CI]) which should be lower than .08 (Byrne, 2010). Values of standardized regression weights equal to or above .50 and squared multiple correlations equal to or above .25 were considered, as well as modification indices (MI > 25) for model adjustment (Marôco, 2010a), although these modifications were post hoc and empirically driven – despite being theoretically justified. Additionally, the Akaike Information Criterion (AIC) was used to compare alternative models, with smaller AIC values indicating superior models and a more stable model for the population under study (Kline, 2016).

Measurement invariance across genders, with five nested models with gradually constricted parameters, were tested in both three- and four-factor models. Model 1 (unconstrained model) tested for configural invariance (basic model structure), Model 2 (measurement weights model) and Model 3 (structural weights model) tested for metric invariance (with the same first, and first and second-order factor loadings across groups, respectively), Model 4 (measurement intercepts model) tested for scalar invariance (constrained factor loadings and item intercepts), and Model 5 (measurement residuals model) tested for strict invariance (same measurement errors). The differences between nested models regarding CFI and RMSEA indices were considered acceptable for the following values: $\Delta CFI \leq 0.02$ and $\Delta RMSEA \leq 0.03$, for tests of metric invariance, and $\Delta CFI \leq 0.01$ and $\Delta RMSEA \leq 0.01$, for tests of scalar invariance (Putnick & Bornstein, 2016).

Scale and item reliability were explored using Cronbach's alpha calculation, with values above .70 considered acceptable (Kline, 2016). Pearson correlations assessed temporal stability and associations with the RTT-A's total scale and factors, and with other relevant constructs. According to Cohen's (1988) standards, correlations from .10 to .29 are considered small, .30 to .49 moderate and ≥ .50 large. Incremental validity of the Test-Irrelevant Thinking dimension was examined using hierarchical multiple regression analyses. The assumptions of normality and homoscedasticity were assessed via inspection of normal probability plots and residuals. Independence of residuals was verified through the Durbin-Watson statistic. The absence of multicollinearity was confirmed by ensuring that Variance Inflation Factor (VIF) values remained below 5, in accordance with recommended guidelines (Marôco, 2010b; Pestana & Gageiro, 2008).

# Results

## Preliminary Analysis

No missing values were found. All values fell within the reference ranges, indicating no significant deviations from normality. Although outliers were detected in 4% of the sample, they were retained in the sample to ensure ecological validity (Hair et al., 2010). Regarding differences in RTT-A's scores by school region (centre vs. south) and type (public vs. private), small effect sizes were found across all comparisons. For school type, students from public schools reported higher test anxiety in the RTT-A total score (whether considering three or four factors) and across all four individual factors (Tension, Worry, Bodily Symptoms, and Test-Irrelevant Thinking), with effect sizes ranging from $d = .18$ to .43. For regional differences, effect sizes ranged from $d = -.02$ to -.25. In this case, students from southern schools reported higher test anxiety on all dimensions except Tension, where no significant difference was observed.

## Confirmatory Factor Analysis

Guided by prior work indicating an overarching test-anxiety construct (Baptista et al., 1989; Sarason, 1984; Vicente & Salvador, 2011), we tested second-order CFA models for a four-factor solution (Tension, Worry, Bodily Symptoms, Test-Irrelevant Thinking) and a three-factor solution (excluding Test-Irrelevant Thinking), using the adolescent item set from Vicente and Salvador (2011). A higher-order latent factor (Test Anxiety) was specified given the substantial correlations among first-order factors (Bollen, 1989).

**Four-Factor Model** The initial model showed suboptimal fit, $\chi^2(523) = 2677.14$, $p < .001$; $\chi^2/df = 5.12$; CFI = .89; TLI = .89; RMSEA = .061, 90% CI [.059, .064]. Several weak loadings emerged for Test-Irrelevant Thinking at the second order ($\lambda = .39$; $R^2 = .16$) and for three items (Worry item 33; Bodily Symptoms items 10 and 26). After removing these items – consistent with prior adaptations (Benson & Bandalos, 1992; Pereira et al., 2021) – and freeing two theoretically coherent residual covariances (Tension 13–22; Worry 23–29), fit improved to acceptable levels, $\chi^2(428) = 2108.68$, $p < .001$; $\chi^2/df = 4.93$; CFI = .91; TLI = .90; RMSEA = .060, 90% CI [.057, .062]. Even so, Test-Irrelevant Thinking retained a weak second-order loading ($\lambda = .37$; $R^2 = .14$).

**Three-Factor Model** The corresponding three-factor model (same adjustments: removal of items 10, 26, 33; residuals 13–22 and 23–29) yielded slightly better fit overall, $\chi^2(204) = 1151.84$, $p < .001$; $\chi^2/df = 5.65$; CFI = .93; TLI = .92;

RMSEA = .065, 90% CI [.062, .069], and a lower AIC (1293.84 vs. 2306.68 for the adjusted four-factor model). Although the normed chi-square remained marginally above conventional cutoffs, this statistic is sensitive to large samples (Byrne, 2010); the pattern for CFI/TLI/RMSEA indicated adequate fit.

In summary, both models achieved acceptable fit after minor, theory-consistent refinements; however, the three-factor solution was marginally superior and more parsimonious, while Test-Irrelevant Thinking consistently showed weak higher-order saturation. All post hoc modifications were empirically indicated and theoretically justified, and should be interpreted with due caution.

## Measurement Invariance for Gender

Measurement invariance analyses provided overall support for configural, metric, and partial scalar invariance of both the three- and four-factor models across gender (Table 1). Changes in CFI and RMSEA were minimal, indicating largely equivalent functioning of RTT-A items for boys and girls. Strict invariance was not fully achieved, as chi-square tests and some CFI differences exceeded conventional thresholds, likely reflecting sample-size sensitivity rather than substantive non-invariance.

Figure 1 depicts both RTT-A models tested in this study. The three-factor model comprises 22 items representing Tension (8 items), Worry (9 items), and Bodily Symptoms (5 items), whereas the four-factor model additionally includes the Test-Irrelevant Thinking dimension (9 items), for a total of 31 items. Item-level means and standard deviations are reported in Table 2.

## Reliability

The RTT-A demonstrated excellent internal consistency for the total score across both the three- and four-factor models. Subscale reliabilities were in the good to excellent range, being highest for Test-Irrelevant Thinking in the four-factor solution and lowest for Bodily Symptoms. Corrected item-total correlations were generally satisfactory across models and slightly higher in the three-factor solution, particularly for Tension items. Item-level diagnostics indicated that removing any item – whether in the 31-item four-factor or the 22-item three-factor version – would not increase internal consistency (see Table 2).

As shown in Table 3, all RTT-A factors correlated strongly with the total score and with one another, except for Test-Irrelevant Thinking, which showed only moderate correlations with the total and the other factors. Temporal stability was high across a five-week retest interval ($n = 120$), with strong correlations between test and retest scores for the total scale and all subscales ($r_{total\_4factors} = .82$, $r_{total\_3factors} = .82$, $r_{tension} = .76$, $r_{worry} = .75$, $r_{test\text{-}irrelevant\ thinking} = .78$, $r_{bodily\_symptoms} = .63$, $p < .01$), supporting the stability of the RTT-A over time.

**Table 1** Measurement invariance across gender (three- and four-factor models)

| | $\chi^2$ | *df* | CFI | RMSEA [90% CI] | $\Delta\chi^2$ | $\Delta df$ | *p* | ΔCFI | ΔRMSEA |
|---|---|---|---|---|---|---|---|---|---|
| ***Three-factor model*** | | | | | | | | | |
| *Summary of fit statistics* | | | | | | | | | |
| Girls | 866.62* | 204 | .903 | .073 [.068/.078] | | | | | |
| Boys | 545.69* | 204 | .932 | .059 [.053/.065] | | | | | |
| *Multi-group analyses* | | | | | | | | | |
| Unconstrained model | 1412.27* | 408 | .916 | .047 [.045/.050] | - | - | - | - | - |
| Measurement weights | 1470.05* | 427 | .912 | .047 [.045/.050] | 57.78 | 19 | < .001 | .004 | .000 |
| Structural weights | 1474.42* | 429 | .912 | .047 [.045/.050] | 4.37 | 2 | .113 | .000 | .000 |
| Measurement intercepts | 1635.25* | 451 | .901 | .049 [.046/.052] | 160.84 | 22 | < .001 | .011 | .002 |
| Measurement residuals | 1758.27* | 475 | .892 | .050 [.047/.052] | 123.02 | 24 | < .001 | .009 | .001 |
| ***Four-factor model*** | | | | | | | | | |
| *Summary of fit statistics* | | | | | | | | | |
| Girls | 1584.32* | 428 | .889 | .067 [.063/.070] | | | | | |
| Boys | 1096.01* | 428 | .914 | .057 [.053/.061] | | | | | |
| *Multi-group analyses* | | | | | | | | | |
| Unconstrained model | 2680.29* | 856 | .900 | .044 [.042/.046] | - | - | - | - | - |
| Measurement weights | 2747.19* | 883 | .897 | .044 [.042/.046] | 66.90 | 27 | < .001 | .003 | .000 |
| Structural weights | 2752.72* | 886 | .897 | .044 [.042/.046] | 5.53 | 3 | .137 | .000 | .000 |
| Measurement intercepts | 2951.12* | 917 | .888 | .045 [.043/.047] | 198.40 | 31 | < .001 | .009 | .001 |
| Measurement residuals | 3080.52* | 950 | .883 | .045 [.046/.047] | 129.41 | 33 | < .001 | .005 | .000 |

Girls: $n = 806$; Boys: $n = 589$. $\chi^2$ = chi-square goodness-of-fit statistic; *df* = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval

*$p < .001$

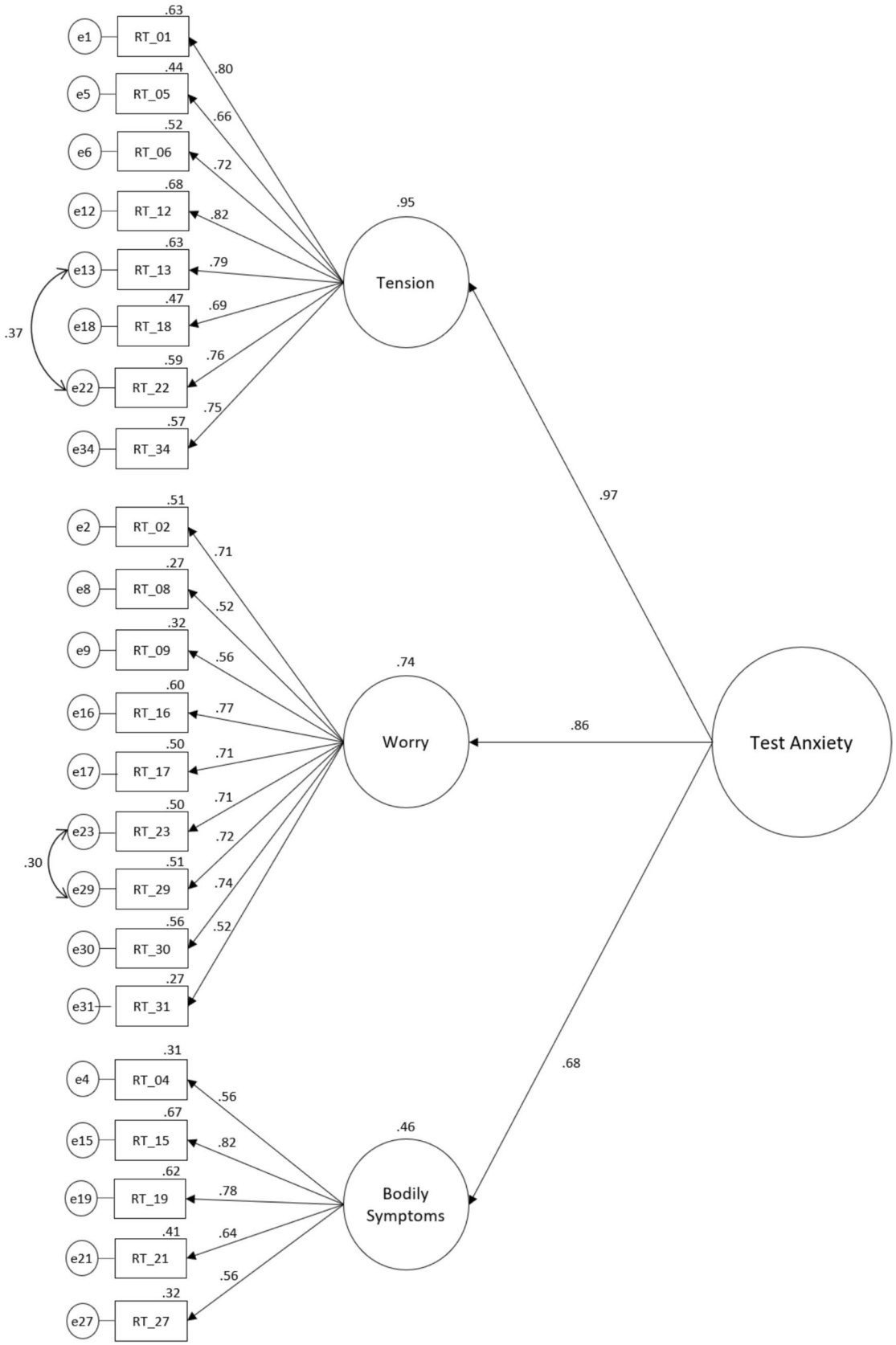


**Fig. 1** CFA models of the RTT-A (three- and four-factor). *Note.* Standardized values are presented. Values above each item indicate standardized multiple correlations. Values above the arrows emerging from each factor represent standardized regression weights. Values next to curved arrows denote correlations between item errors

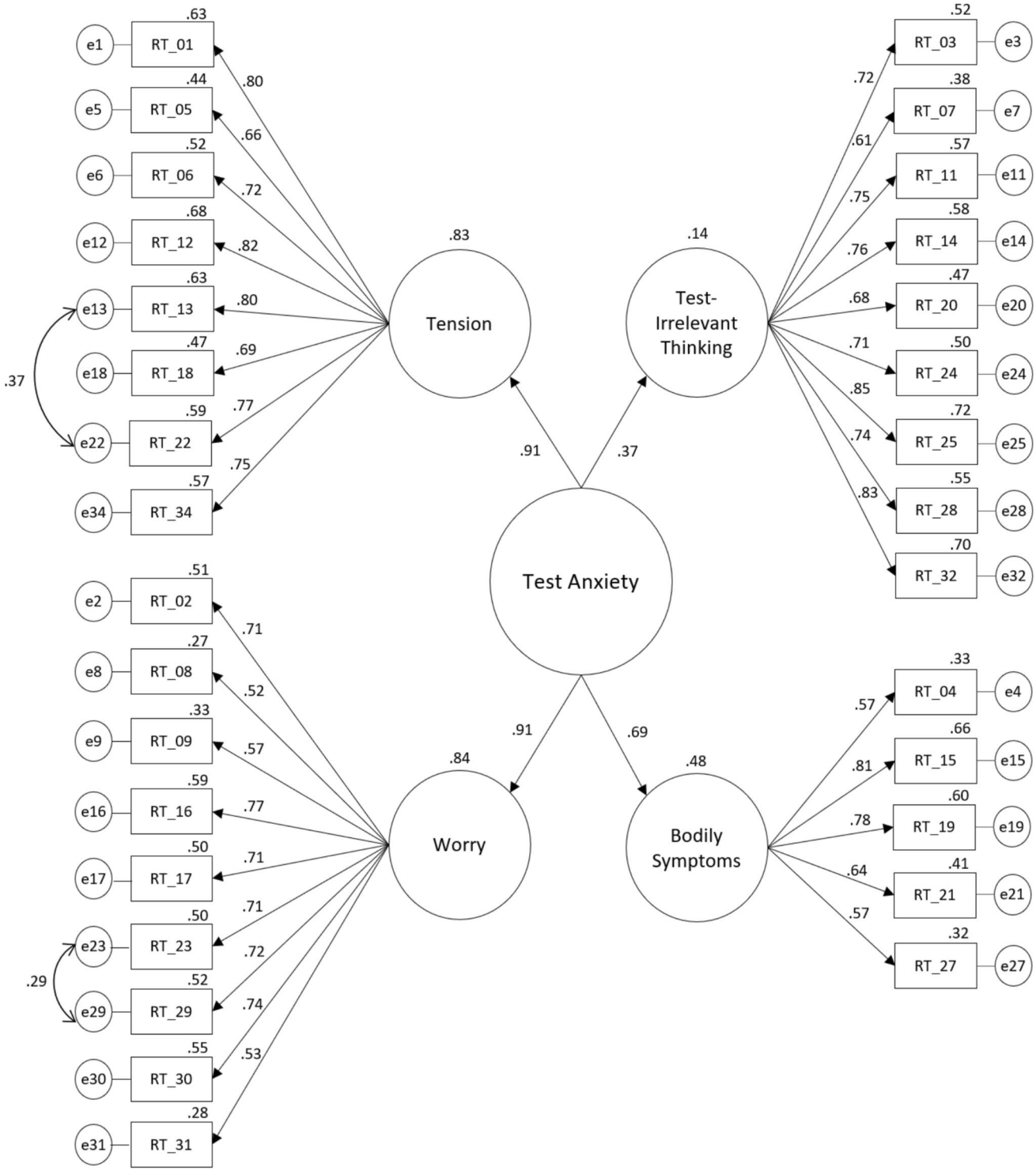


**Fig. 1** (continued)

**Table 2** Item-level descriptive statistics and reliability for the RTT-A

| | α total (3 and 4 factors) = .937 | | | | | |
|---|---|---|---|---|---|---|
| Items | *M (SD)* | Item-total *r* (four-factor model) | Item-total *r* (three-factor model) | α | α if item deleted (four-factor model) | α if item deleted (three-factor model) |
| *Tension* | – | – | – | **.912** | – | – |
| 1. I feel distressed and uneasy before tests | 2.25 (0.93) | .65 | .73 | – | .934 | .932 |
| 5. I freeze up when I think about an upcoming test | 1.73 (0.83) | .63 | .66 | – | .934 | .934 |
| 6. I feel jittery before tests | 1.90 (0.95) | .59 | .66 | – | .934 | .934 |
| 12. While taking a test, I feel tense | 2.45 (1.00) | .66 | .75 | – | .933 | .932 |
| 13. I find myself becoming anxious the day of a test | 2.55 (1.03) | .60 | .71 | – | .934 | .933 |
| 18. I wish tests did not bother me so much | 2.82 (1.14) | .59 | .65 | – | .934 | .934 |
| 22. I am anxious about tests | 2.44 (1.04) | .58 | .69 | – | .934 | .933 |
| 34. I have an uneasy feeling before an important test | 2.12 (1.11) | .60 | .69 | – | .934 | .933 |
| *Worry* | – | – | – | **.878** | – | – |
| 2. The thought "What happens if I fail this test?" goes through my mind during tests | 2.51 (0.98) | .61 | .65 | – | .934 | .934 |
| 8. During a difficult test, I worry whether I will pass it | 2.84 (1.01) | .41 | .44 | – | .936 | .937 |
| 9. While taking a test, I find myself thinking how much brighter the other people are | 1.90 (1.02) | .58 | .53 | – | .934 | .936 |
| 16. Before taking a test, I worry about failure | 2.59 (0.94) | .61 | .68 | – | .934 | .933 |
| 17. While taking a test, I often think about how difficult it is | 2.59 (0.93) | .61 | .63 | – | .934 | .934 |
| 23. Thoughts of doing poorly interfere with my concentration during tests | 2.65 (0.96) | .65 | .66 | – | .934 | .933 |
| 29. During tests, I think about how poorly I am doing | 2.42 (0.93) | .57 | .66 | – | .933 | .934 |
| 30. Before tests, I feel troubled about what is going to happen | 2.53 (0.97) | .62 | .68 | – | .934 | .933 |
| 31. The harder I work at taking a test, the more confused I get | 1.85 (0.86) | .57 | .52 | – | .934 | .936 |
| *Test-irrelevant thinking* | – | – | – | **.916** | – | – |
| 3. During tests, I find myself thinking of things unrelated to the material being tested | 2.65 (0.92) | .40 | – | – | .936 | – |
| 7. Irrelevant bits of information pop into my head during a test | 2.49 (0.87) | .52 | – | – | .935 | – |
| 11. My mind wanders during tests | 2.34 (0.95) | .54 | – | – | .935 | – |
| 14. I think about current events during a test | 2.10 (0.96) | .50 | – | – | .935 | – |
| 20. I have fantasies a few times during a test | 1.89 (1.02) | .41 | – | – | .936 | – |
| 24. While taking tests, I sometimes think about being somewhere else | 2.06 (1.02) | .51 | – | – | .935 | – |
| 25. During tests, I find I am distracted by thoughts of upcoming events | 2.29 (0.99) | .51 | – | – | .935 | – |
| 28. I daydream during tests | 1.83 (0.99) | .44 | – | – | .936 | – |
| 32. During tests I think about recent past events | 2.11 (0.98) | .52 | – | – | .935 | – |
| *Bodily symptoms* | – | – | – | **.801** | – | – |
| 4. I become aware of my body during tests (feeling itches, pain, sweat, nausea) | 1.77 (0.83) | .53 | .48 | – | .935 | .936 |
| 15. I get a headache during an important test | 1.64 (0.86) | .55 | .57 | – | .935 | .935 |
| 19. I get a headache before a test | 1.51 (0.81) | .51 | .54 | – | .935 | .935 |
| 21. I sometimes feel dizzy after a test | 1.47 (0.79) | .43 | .44 | – | .936 | .937 |
| 27. My mouth feels dry during a test | 1.70 (0.89) | .48 | .49 | – | .935 | .936 |

Table includes item means, standard deviations, item–total correlations, Cronbach's alpha for each factor, and Cronbach's alpha if item deleted, for both the three- and four-factor models

## Relationship to Other Variables

Correlations between the RTT-A and other anxiety measures were examined in a subsample of 299 adolescents (Table 4). The RTT-A total and its core factors – particularly Tension and Bodily Symptoms – showed the strongest associations with physical symptoms of anxiety. Performance anxiety in formal social situations and general social anxiety were also moderately related to the RTT-A, whereas Test-Irrelevant Thinking correlated only weakly with these domains. In contrast, correlations with separation anxiety and harm avoidance were small. Notably, Test-Irrelevant Thinking showed negligible associations

**Table 3** Correlations among RTT-A factors and total score (three- and four-factor models)

| | RTT-A Total (four-factor model) | RTT-A Total (three-factor model) | Tension | Worry | Test-Irrelevant Thinking | Bodily Symptoms |
|---|---|---|---|---|---|---|
| Tension | .82* | .93* | – | | | |
| Worry | .85* | .89* | .72* | – | | |
| Test-Irrelevant Thinking | .69* | .37* | .25* | .42* | – | |
| Bodily Symptoms | .68* | .71* | .57* | .59* | .33* | – |

* $p < .01$

**Table 4** RTT-A correlations with related measures (three- and four-factor models)

| | SAASA_PFSS | MASC_PS | MASC_SocA | MASC_SepA | MASC_HA |
|---|---|---|---|---|---|
| RTT-A Total (four-factor model) | .46** | .57** | .46** | .29** | .29** |
| RTT-A Total (three-factor model) | .48** | .56** | .48** | .30** | .35** |
| RTT-A Tension | .41** | .51** | .41** | .28** | .33** |
| RTT-A Worry | .46** | .44** | .46** | .27** | .32** |
| RTT-A Test-Irrelevant Thinking | .20** | .33** | .20** | .12* | .05 |
| RTT-A Bodily Symptoms | .34** | .50** | .35** | .22** | .27** |

*SAASA* Social Anxiety and Avoidance Scale for Adolescents; *SAASA_PFSS* Performance in formal social situations factor of the discomfort/anxiety subscale of the SAASA; *MASC* Multidimensional Anxiety Scale for Children; *MASC_PS* Physical symptoms factor of the MASC; *MASC_SocA* Social Anxiety factor of the MASC; *MASC_SepA* Separation anxiety factor of the MASC; *MASC_HA* Harm avoidance factor of the MASC

* $p < .05$; ** $p < .01$

with harm avoidance, whereas Tension correlated moderately with this construct. Overall, the RTT-A demonstrated a coherent pattern of stronger relationships with conceptually proximal anxiety forms and weaker associations with distal domains, confirming its convergent and divergent validity.

## Incremental Validity

To assess the incremental validity of the Test-Irrelevant Thinking dimension, hierarchical regression analyses were conducted using two criterion variables: anxiety specific to performance in formal social situations (SAASA) and social anxiety (MASC). In both models, the RTT dimensions were entered sequentially: Tension (Step 1), Worry (Step 2), Bodily Symptoms (Step 3), and Test-Irrelevant Thinking (Step 4), to determine whether the latter added explanatory value beyond the core components.

For performance in formal social situations (SAASA), the model with Tension alone was significant, $F(1, 297) = 61.64$, $p < .001$, explaining 17.2% of the variance ($\beta = .42$, $p < .001$). Adding Worry significantly improved the model, $\Delta R^2 = .059$, $F$ change $(1, 296) = 22.74$, $p < .001$, with both Worry ($\beta = .35$, $p < .001$) and Tension ($\beta = .16$, $p = .026$) emerging as significant predictors. The inclusion of Bodily Symptoms further improved the model, $\Delta R^2 = .014$, $F$ change $(1, 295) = 5.59$, $p = .019$; Bodily Symptoms ($\beta = .15$, $p = .019$) and Worry ($\beta = .33$, $p < .001$) remained significant, whereas Tension lost significance ($\beta = .09$, $p = .239$). In the final step, the inclusion of Test-Irrelevant Thinking did not improve the model, $\Delta R^2 = .000$, $F$ change $(1, 294) = 0.004$, $p = .948$, and it was not a significant predictor ($\beta = .004$, $p = .948$). Worry and Bodily Symptoms remained the only significant predictors.

A similar pattern was found for social anxiety (MASC). The model including only Tension was significant, $F(1, 297) = 61.44$, $p < .001$, accounting for 17.1% of the variance ($\beta = .41$, $p < .001$). Adding Worry significantly improved the model ($\Delta R^2 = .054$, $p < .001$), with both Worry ($\beta = .33$, $p < .001$) and Tension ($\beta = .18$, $p = .018$) as significant predictors. In the third step, Bodily Symptoms added further explanatory power ($\Delta R^2 = .013$, $p = .028$), remaining significant ($\beta = .14$, $p = .028$) alongside Worry ($\beta = .32$, $p < .001$), while Tension became non-significant ($\beta = .11$, $p = .172$). The final step, which added Test-Irrelevant Thinking, did not improve the model ($\Delta R^2 = .000$, $p = .866$), and this dimension was not a significant predictor ($\beta = .01$, $p = .866$). Worry and Bodily Symptoms remained the only significant predictors.

## Descriptive Data

Descriptive statistics for the RTT-A total and subscale scores are presented in Table 5. Girls reported significantly higher levels of test anxiety than boys on the total scale (in both the

**Table 5** Descriptive statistics and gender differences for the RTT-A (three- and four-factor models)

| | M (SD) | | | *Mean differences* | |
|---|---|---|---|---|---|
| | Total | Boys | Girls | *t* | *d* |
| | *(N = 1395)* | *(n = 589)* | *(n = 806)* | *df = 1393* | |
| RTT-A Total (four-factor model) | 65.88 (15.56) | 62.80 (15.27) | 68.00 (15.45) | −3.88** | .34 |
| RTT-A Total (three-factor model) | 48.34 (13.65) | 43.75 (12.60) | 52.01 (13.35) | −4.03*** | .64 |
| Tension | 20.44 (6.27) | 18.65 (6.20) | 21.68 (6.04) | −4.23*** | .50 |
| Worry | 21.31 (5.67) | 19.95 (5.69) | 22.24 (5.48) | −3.50*** | .41 |
| Test-Irrelevant Thinking | 19.01 (6.20) | 19.25 (6.26) | 18.84 (6.18) | .57 | .07 |
| Bodily Symptoms | 7.64 (2.75) | 7.22 (2.48) | 7.92 (2.90) | −2.17* | .26 |

*t* = t test statistic; *df* = degrees of freedom; *d* = Cohen's *d* for effect size
*** $p < .001$; ** $p < .01$, * $p < .05$

three- and four-factor models) and across most subscales. The only exception was Test-Irrelevant Thinking, for which boys scored slightly higher, though this difference was nonsignificant. Effect sizes were generally small (Cohen's $d \approx .26$-.41), with Tension showing the largest difference ($d = .50$) and the total score of the three-factor model reaching a medium effect size ($d = .64$).

## Discussion

This study sought to validate the RTT-A in a large sample of Portuguese adolescents, addressing the lack of multidimensional test-anxiety measures with cross-cultural validity and broad international recognition within the test-anxiety literature for this population. Building on previous exploratory work (Vicente & Salvador, 2011), we examined the factor structure, measurement invariance across gender, psychometric properties, associations with related anxiety constructs, and gender differences. Importantly, we also compared two competing models – one including the Test-Irrelevant Thinking factor and another excluding it – to examine whether the inclusion of this dimension improves model fit or predictive utility. In doing so, this study provides empirical support for the RTT-A as a psychometrically robust measure of test anxiety in Portuguese adolescents and offers new insight into the debated role of Test-Irrelevant Thinking within this framework. Overall, both the three- and four-factor models demonstrated adequate psychometric properties, including good internal consistency, test–retest reliability, and evidence of convergent and divergent validity. Measurement invariance across gender was also supported.

Findings were consistent with Hypothesis 1 – confirming that the four-factor model achieved acceptable fit indices – although the three-factor solution (excluding Test-Irrelevant Thinking) demonstrated slightly superior fit and lower AIC values. To further improve model fit, three items – items 10 and 26 from Bodily Symptoms and item 33 from Worry – were removed due to low statistical weight and conceptual ambiguity. Item 33 (*During tests, I wonder how the other people are doing*) overlapped conceptually with both worry and distraction, whereas items 10 (*I feel the need to go to the toilet more often than usual during a test*) and 26 (*My hands often feel cold before and during a test*) reflected somatic experiences that, although plausible, may be less prevalent or more idiosyncratic among adolescents. These items also showed weak loadings in prior RTT adaptations (Benson & Bandalos, 1992; Pereira et al., 2021). Additionally, modification indices suggested correlating residuals between two pairs of semantically overlapping items – 13–22 within Tension (*I find myself becoming anxious the day of a test*; *I am anxious about tests*) and 23–29 within Worry (*Thoughts of doing poorly interfere with my concentration during tests*; *During tests, I think about how poorly I am doing*) – which improved model fit and were theoretically justified. Such post hoc adjustments are standard in psychometric refinement but, having been applied within the same dataset, may have introduced overfitting. Consequently, the refined model should be interpreted cautiously and replicated in independent samples to confirm its robustness.

Evidence for both the three- and four-factor models was supported through configural and metric invariance across gender, confirming the equivalence of factor structure and factor loadings (Hypothesis 2). Scalar invariance was partial rather than full, whereas strict invariance was not fully achieved; however, most fit indices met accepted thresholds, suggesting that these deviations likely reflect sample-size sensitivity rather than substantive non-invariance (Counsell et al., 2020). This pattern suggests largely comparable functioning of RTT-A items for boys and girls, although mean-level gender comparisons should be interpreted with some caution. These results are consistent with findings from the Portuguese short version of the RTT (Pereira et al., 2021) and add to preliminary support for two gender invariant versions of the RTT-A in Portuguese adolescents: a three-factor version with 22 items, comprising Tension (8 items), Worry (9 items), and Bodily Symptoms (5 items), and a four-factor version, which includes the same items in those three factors plus Test-Irrelevant Thinking (9 items), for a total of 31 items. The three-factor version, however, appears slightly more parsimonious and psychometrically robust in this sample.

Reliability analyses (Hypothesis 3) indicated high internal consistency for the total scale and good to excellent reliability for all subscales. Tension and Test-Irrelevant Thinking showed the highest α values, whereas Bodily Symptoms was somewhat lower but still within acceptable thresholds – consistent with previous work (Pereira et al., 2021; Roos et al., 2021). The comparatively lower reliability of Bodily Symptoms may reflect the greater heterogeneity and situational variability of somatic manifestations in evaluative contexts, which are often less cognitively coherent than worry-based components (Chawla et al., 2025; Overmeyer & Endrass, 2023; Roos et al., 2021, 2022). Importantly, strong temporal stability across all factors suggests that the RTT-A captures relatively stable individual differences in adolescents' responses to evaluative situations, rather than transient state fluctuations. Together, these reliability patterns support the use of the RTT-A as a psychometrically robust instrument for assessing test anxiety dimensions in adolescent samples.

Convergent and divergent validity findings (Hypothesis 4) aligned with theoretical expectations. Total RTT-A scores – across both the three- and four-factor models – showed moderate to strong associations with constructs conceptually related to test anxiety, namely physical symptoms, social anxiety, and performance anxiety in formal social situations. The strongest correlations emerged between Tension and Bodily Symptoms and the MASC physical-symptoms subscale, highlighting the RTT-A's sensitivity to the affective and physiological manifestations of evaluative stress, such as autonomic arousal and somatic tension (Roos et al., 2021, 2022; Sarason, 1984; Zeidner, 1998). Moderate correlations with social-evaluative domains – social anxiety and performance anxiety in formal social situations – further supported convergent validity while indicating only partial overlap, suggesting that the RTT-A captures anxiety responses specific to evaluative threat rather than general social discomfort (Kavanagh et al., 2017; Knappe et al., 2011; Laurin-Barantke et al., 2016). Conversely, total RTT-A scores showed only small associations with separation anxiety and harm avoidance – both considered more distal anxiety domains (Hasani et al., 2025; Putwain et al., 2021a, 2021b) – thus supporting the measure's conceptual specificity. A few exceptions were observed, including moderate correlations between Tension and harm avoidance, and between the three-factor total score and both separation anxiety and harm avoidance. These slightly stronger associations may reflect shared features such as perfectionistic tendencies or heightened sensitivity to authority figures, both of which are known to co-occur with test-related stress (Putwain et al., 2016b, 2017).

Finally, Hypothesis 5 was also supported: girls scored higher than boys on the total RTT-A and on the Tension, Worry, and Bodily Symptoms subscales, with small-to-moderate effect sizes, similar to previous studies (Núñez-Peña et al., 2016; Putwain & Daly, 2014). No gender differences emerged for Test-Irrelevant Thinking, which slightly reduced the overall gender effect when included in the total score. These results replicate prior research indicating that gender differences in test anxiety are modest and partly reflect psychosocial and cognitive factors, such as internalized academic expectations and fear of failure (Ben-Zur & Zeidner, 2012; di Maria & di Nuovo, 1990; Egloff & Schmukle, 2004; Núñez-Peña et al., 2016; Zeidner, 1998). In line with this modest magnitude, and considering that scalar invariance across gender was partial, these mean-level differences should be interpreted with some caution, as a small proportion of items may function differently for boys and girls.

Given the consistent pattern observed across analyses, the role of the Test-Irrelevant Thinking factor deserves particular consideration. Across all models and validity tests, this dimension displayed weaker factor loadings, modest correlations with the other RTT-A factors, and no incremental predictive value once core components such as Worry and Bodily Symptoms were accounted for. These results, together with its minimal associations with both proximal (social and performance anxiety) and distal (separation anxiety and harm avoidance) constructs, suggest that Test-Irrelevant Thinking may not reflect a direct manifestation of anxiety. Instead, it appears to capture a peripheral cognitive process – possibly attentional disengagement or cognitive interference that occurs during evaluative tasks. Theoretically, this aligns with models of unintentional mind-wandering and task-unrelated thought (Figueiredo et al., 2020; Seli et al., 2016; Smallwood & Schooler, 2014), which describe fluctuations of attention away from task-relevant stimuli under conditions of cognitive strain or emotional load. Unlike Worry, which entails anxious preoccupation, Test-Irrelevant Thinking may index automatic shifts of attention or avoidance of self-referential threat cues – phenomena adjacent to, but not constitutive of, test anxiety.

This interpretation is further supported by Test-Irrelevant Thinking's comparatively low intercorrelations with Tension and Worry and by its independence from gender effects, in contrast to the small-to-moderate gender differences observed for the other RTT-A dimensions. Together, these findings reinforce the notion that Test-Irrelevant Thinking may reflect a distinct cognitive mechanism rather than an affective response to evaluative pressure. In this sense, it may represent an attentional vulnerability that co-occurs with test anxiety but does not define it. This distinction may partly relate to underlying neurocognitive profiles, as inattentive symptoms have been associated with heightened cognitive interference during evaluative tasks (Arnold

et al., 2015; Owens et al., 2014). Although the RTT-A was not designed to assess attention-related traits, such evidence suggests that Test-Irrelevant Thinking may capture non-anxious attentional lapses (e.g., unintentional mind-wandering) rather than emotional reactivity, highlighting the need to further examine its cognitive underpinnings. Nevertheless, given that several previous studies retained this factor despite similar patterns of weak loadings (Baptista et al., 1989; Benson & Bandalos, 1992; Pereira et al., 2021; Vicente & Salvador, 2011), and considering potential developmental or contextual variability, we do not recommend its definitive exclusion from the RTT framework. Instead, its inclusion should depend on research aims: for studies focused on cognitive interference, attentional lapses, or mind-wandering under evaluative pressure, the four-factor model remains informative; however, for general assessments of test anxiety in adolescents, the three-factor model offers a more parsimonious, theoretically coherent, and empirically supported structure.

### Limitations and Future Directions

This study presents several limitations that warrant consideration. First, although the sample was large and included both public and private schools, it was geographically limited to central and southern Portugal due to logistical constraints. This restricts representativeness, particularly regarding adolescents from northern or rural areas. Moreover, the Portuguese educational context – marked by high-stakes national assessments during secondary education – may intensify test anxiety in culturally specific ways, limiting cross-cultural generalizability. Future research should employ nationally representative samples and examine the RTT-A in diverse cultural and educational contexts to assess its robustness and applicability across populations. Second, all data were self-reported, which may introduce biases such as social desirability or distorted recall, particularly in constructs involving introspective processes (Althubaiti, 2016). Multi-method designs incorporating behavioral, observational, or informant-based measures could mitigate these limitations and provide a more comprehensive assessment of test anxiety.

A central methodological limitation concerns the analytic process used to refine the model. Item removal and the addition of correlated residuals were post hoc, empirically driven decisions applied within the same sample, increasing the risk of overfitting. While theoretically informed and common in psychometric research (Thompson, 2004), such adjustments compromise the confirmatory status of the model. Accordingly, we frame the final solutions as exploratory and stress the need for replication in independent and longitudinal samples.

A particularly salient issue concerns the Test-Irrelevant Thinking factor. Importantly, our data does not allow for a definitive conclusion regarding the superiority of the three- versus four-factor model. Both versions demonstrated a good fit, and while the three-factor model offered slightly better parsimony and AIC values, these differences were not statistically decisive. We thus caution against overinterpreting the superiority of either structure and highlight the need for replication using independent samples and alternative outcome variables. Additionally, the absence of diagnostic or neurocognitive data (e.g., ADHD symptoms, learning difficulties) limited our ability to explore whether Test-Irrelevant Thinking reflects comorbid attentional traits that exacerbate test anxiety without being intrinsic to it. Future studies should assess Test-Irrelevant Thinking in relation to executive functioning, academic performance, and validated mind-wandering measures (e.g., Mowlem et al., 2019; Mrazek et al., 2013), and examine its potential value as a cognitive avoidance strategy rather than an anxiety dimension per se.

Therefore, we refrain from definitively recommending the exclusion of Test-Irrelevant Thinking from the RTT across all contexts. Instead, along with further validation, we suggest that its inclusion should be contingent on study aims: for clinical or research contexts interested in attentional disengagement, Test-Irrelevant Thinking may offer unique value; for general assessments of test anxiety, the three-factor model may be more efficient and conceptually coherent. Finally, longitudinal research is needed to track how test anxiety and its components evolve over time and under academic stress, and to distinguish stable traits from state-dependent responses. Such designs would also inform whether the role of Test-Irrelevant Thinking changes with developmental maturation, coping strategies, or intervention exposure.

## Conclusion

This study provides empirical support for the RTT-A as a multidimensional instrument for assessing test anxiety in Portuguese adolescents. Both the three- and four-factor versions demonstrated good psychometric properties, including good reliability, temporal stability, factorial invariance across gender, and evidence of convergent and divergent validity. However, findings consistently questioned the contribution of the Test-Irrelevant Thinking dimension – both in terms of structural coherence and predictive utility.

The three-factor version of the RTT-A, excluding Test-Irrelevant Thinking, may offer a more parsimonious and theoretically grounded model of test anxiety in adolescence, capturing its primary affective-physiological (Tension,

Bodily Symptoms) and cognitive (Worry) components. Still, its inclusion may be appropriate in studies targeting attentional disengagement, whereas the three-factor model may be preferable for general assessment purposes. The RTT-A, particularly in its more parsimonious form, offers researchers and practitioners a concise and theoretically grounded tool to assess test anxiety in adolescents, supporting screening efforts, informing intervention planning, and guiding education policy decisions related to academic stress and evaluation. Its cross-cultural validation potential makes it especially useful for international comparisons and for tailoring interventions to specific developmental and cultural contexts. However, future work must rigorously test its structural stability and clinical relevance across time, settings, and populations – especially as we seek to understand how adolescents experience, manage, and overcome evaluative stress.

**Acknowledgments** We express our sincere gratitude to all the adolescents who voluntarily participated in this study, as well as to their parents/legal guardians for their consent. We also extend our deep appreciation to Professor Irwin G. Sarason for granting permission to adapt the RTT for use with Portuguese adolescents.

**Author Contribution** CPP: conceptualization, data curation, formal analysis, investigation, methodology, writing – original draft. DWP: supervision, writing – review and editing. SGH: supervision, writing – review and editing. ARV: investigation. MSC: conceptualization, methodology, supervision, writing – review and editing. All authors approved the final version of the manuscript.

**Funding** Open access funding provided by FCT|FCCN (b-on). This work was funded by Fundação para a Ciência e Tecnologia(FCT), under HEI-Lab R&D Unit (UIDB/05380/2020, https://doi.org/10.54499/UIDB/05380/2020). Fundação para a Ciência e a Tecnologia, SFRH/BD/143520/2019,Cláudia P. Pires

**Data Availability** The data that support the findings of this study are available on request from the corresponding author, CPP. The data arenot publicly available due to their containing information that could compromise the privacy of research participants.

## Declarations

**Ethics Statement** This study was approved by the Ethics Committee of the Faculty of Psychology and Educational Sciences of the University of Coimbra (CEDI28.11.2019) and has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

**Conflict of Interest** The authors declare that they have no conflict ofinterest.



## References

Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare, 9*(1), 211–217. https://doi.org/10.2147/JMDH.S104807

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. American Psychiatric Association.

American Psychiatric Association (2022). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., text rev.). American Psychiatric Association.

Ang, R. P., Klassen, R. M., Chong, W. H., Huan, V. S., Wong, I. Y., Yeo, L. S., & Krawchuk, L. L. (2009). Cross-cultural invariance of the Academic Expectations Stress Inventory: Adolescent samples from Canada and Singapore. *Journal of Adolescence, 32*(5), 1225–1237. https://doi.org/10.1016/j.adolescence.2009.01.009

Arbuckle, J. L. (2017). *Amos* (Version 25.0). [Computer program]. IBM SPSS.

Arnold, L. E., Hodgkins, P., Kahle, J., Madhoo, M., & Kewley, G. (2015). Long-term outcomes of ADHD: Academic achievement and performance. *Journal of Attention Disorders, 24*(1), 73–85. https://doi.org/10.1177/1087054714566076

Banks, J., & Smyth, E. (2015). 'Your whole life depends on it': Academic stress and high-stakes testing in Ireland. *Journal of Youth Studies, 18*(5), 598–616. https://doi.org/10.1080/13676261.2014.992317

Baptista, A., Soczka, L., & Pinto, A. (1989). Ansiedade dos exames: Aplicação de reacções aos testes (RT) a uma amostra da população portuguesa [Exam anxiety: Application of the reactions to tests (RT) to a sample of the portuguese population]. *Psicologia, 7*(1), 39–49. https://doi.org/10.17575/rpsicol.v7i1.754

Benson, J., & Bandalos, D. L. (1992). Second-order confirmatory factor analysis of the Reactions to Tests scale with cross-validation. *Multivariate Behavioral Research, 27*(3), 459–487. https://doi.org/10.1207/s15327906mbr2703_8

Ben-Zur, H., & Zeidner, M. (2012). Gender differences in loss of psychological resources following experimentally-induced vicarious stress. *Anxiety, Stress & Coping, 25*(4), 457–475. https://doi.org/10.1080/10615806.2011.619526

Bögels, S. M., Alden, L., Beidel, D. C., Clark, L. A., Pine, D. S., Stein, M. B., & Voncken, M. (2010). Social anxiety disorder: Questions and answers for the DSM-V. *Depression and Anxiety, 27*(2), 168–189. https://doi.org/10.1002/da.20670

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Bound, J., Hershbein, B., & Terry Long, B. (2009). Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives, 23*(4), 119–146. https://doi.org/10.1257/jep.23.4.119

Byrne, B. M. (2010). *Structural equation modelling with Amos: Basic concepts, applications, and programming* (2nd ed.). Taylor and Francis Group.

Cassady, J. C., & Finch, W. H. (2015). Using factor mixture modeling to identify dimensions of cognitive test anxiety. *Learning and*

*Individual Differences, 41*, 14–20. https://doi.org/10.1016/j.lindif.2015.06.002

Chamberlain, S., Daly, A. L., & Spalding, V. (2011). The fear factor: Students' experiences of test anxiety when taking A-level examinations. *Pastoral Care in Education, 29*(3), 193–205. https://doi.org/10.1080/02643944.2011.599856

Chawla, J. K., Bala, I., Dhiman, C., Dixit, P. V., Paliwal, V., Bhattacharjee, B., & Sharma, S. K. (2025). Psychological analysis of anxiety and its impact on athletic performance in competitive sports. *Retos, 72*, 1174–1185. https://doi.org/10.47197/retos.v71.117070

Chou, M. (2018). Predicting self-efficacy in test preparation: Gender, value, anxiety, test performance, and strategies. *Journal of Educational Research, 112*(1), 61–71. https://doi.org/10.1080/00220671.2018.1437530

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Conner, J., Pope, D., & Galloway, M. (2009). Success with less stress. *Educational Leadership, 67*(4), 54–57.

Counsell, A., Cribbie, R. A., & Flora, D. B. (2020). Evaluating equivalence testing methods for measurement invariance. *Multivariate Behavioral Research, 55*(2), 312–328. https://doi.org/10.1080/00273171.2019.1633617

Cunha, M., & Paiva, M. J. (2012). Text anxiety in adolescents: The role of self-criticism and acceptance and mindfulness skills. *The Spanish Journal of Psychology, 15*(2), 533–543. https://doi.org/10.5209/REV_SJOP.2012.V15.N2.38864

Cunha, M., Pinto-Gouveia, J., Salvador, MdoC., & Alegre, S. (2004). Medos sociais na adolescência: A Escala de Ansiedade e Evitamento de Situações Sociais para Adolescentes (EAESSA) [Social fears in adolescence: The social situations anxiety and avoidance scale for adolescents (SSAASA)]. *Psychologica, 36*, 195–217.

Cunha, M., Pinto-Gouveia, J., & Salvador, MdoC. (2008). Social fears in adolescence: The social anxiety and avoidance scale for adolescents. *European Psychologist, 13*(3), 197–213. https://doi.org/10.1027/1016-9040.13.3.197

D'Agati, E., Curatolo, P., & Mazzone, L. (2019). Comorbidity between ADHD and anxiety disorders across the lifespan. *International Journal of Psychiatry in Clinical Practice, 23*(4), 238–244. https://doi.org/10.1080/13651501.2019.1628277

di Maria, F., & di Nuovo, S. (1990). Gender differences in social and test anxiety. *Personality and Individual Differences, 11*(5), 525–530. https://doi.org/10.1016/0191-8869(90)90066-Z

Egloff, B., & Schmukle, S. C. (2004). Gender differences in implicit and explicit anxiety measures. *Personality and Individual Differences, 36*(8), 1807–1815. https://doi.org/10.1016/J.PAID.2003.07.002

Feld, L. D., & Shusterman, A. (2015). Into the pressure cooker: Student stress in college preparatory high schools. *Journal of Adolescence, 41*, 31–42. https://doi.org/10.1016/j.adolescence.2015.02.003

Field, A. (2009). *Discovering Statistics Using IBM SPSS Statistics* (3rd ed.). SAGE Publications.

Figueiredo, T., Lima, G., Erthal, P., Martins, R., Corção, P., Leonel, M., Ayrão, V., Fortes, D., & Mattos, P. (2020). Mind-wandering, depression, anxiety and ADHD: Disentangling the relationship. *Psychiatry Research, 285*, Article 112798. https://doi.org/10.1016/j.psychres.2020.112798

Fisher, D. G., Hageman, A. M., & West, A. N. (2023). Academic burnout among accounting majors: The roles of self-compassion, test anxiety, and maladaptive perfectionism. *Accounting Education, 33*(6), 791–815. https://doi.org/10.1080/09639284.2023.2257672

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

Gerwing, T. G., Rash, J. A., Allen Gerwing, A. M., Bramble, B., & Landine, J. (2015). Perceptions and incidence of test anxiety. *The Canadian Journal for the Scholarship of Teaching and Learning*, 6(3). https://doi.org/10.5206/cjsotl-rcacea.2015.3.3

Gibbons, J. L., & Poelker, K. E. (2019). Adolescent development in a cross-cultural perspective. In K. D. Keith (Ed.), *Cross-cultural psychology: Contemporary themes and perspectives* (2nd ed., pp. 190–215). Wiley Blackwell. https://doi.org/10.1002/9781119519348.ch9

Grant, K. E., McMahon, S. D., Carter, J. S., Carleton, R. A., Adam, E. K., & Chen, E. (2014). The influence of stressors on the development of psychopathology. In M. Lewis & K. D. Rudolph (Eds.), *Handbook of Developmental Psychopathology* (3rd ed., pp. 205–223). Springer. https://doi.org/10.1007/978-1-4614-9608-3_11

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Hasani, H., Manoochehri, M., Zarezadeh, F., Ramezanian, F., Bazghaleh, M., Alafchi, B., & Chekeni, A. M. (2025). Internet addiction, test anxiety, adult separation anxiety disorder: Is there a connection? *Digital Health*. https://doi.org/10.1177/20552076251361678

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

IBM Corp. (2017). *IBM SPSS statistics for windows (Version 25.0) [Computer program]*. IBM Corp.

Johnson, S. B., Blum, R. W., & Giedd, J. N. (2009). Adolescent maturity and the brain: The promise and pitfalls of neuroscience research in adolescent health policy. *Journal of Adolescent Health, 45*(3), 216–221. https://doi.org/10.1016/J.JADOHEALTH.2009.05.016

Kavanagh, B. E., Harvey, J. T., & Mesagno, C. (2017). Social anxiety mediates the relationship between social connectedness and test anxiety: An exploratory investigation. *Journal of Theoretical Social Psychology, 1*(2), 60–69. https://doi.org/10.1002/JTS5.14

Kline, R. B. (2016). *Methodology in the social sciences. Principles and practice of structural equation modelling* (4th ed.). Guilford Press.

Knappe, S., Beesdo-Baum, K., Fehm, L., Stein, M. B., Lieb, R., & Wittchen, H.-U. (2011). Social fear and social phobia types among community youth: Differential clinical features and vulnerability factors. *Journal of Psychiatric Research, 45*(1), 111–120. https://doi.org/10.1016/j.jpsychires.2010.05.002

Laurin-Barantke, L., Hoyer, J., Fehm, L., & Knappe, S. (2016). Oral but not written test anxiety is related to social anxiety. *World Journal of Psychiatry, 6*(3), 351–357. https://doi.org/10.5498/WJP.V6.I3.351

Levesque, R. J. R. (2018). Achievement tests. In R. J. R. Levesque (Ed.), *Encyclopedia of Adolescence* (2nd ed., pp. 64–66). Springer.

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

Lowe, P. A. (2019). Exploring cross-cultural and gender differences in test anxiety among U. S. and Canadian college students. *Journal of Psychoeducational Assessment, 37*(1), 12–118. https://doi.org/10.1177/0734282917724904

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

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*(6), 503–514. https://doi.org/10.1177/0734282910395894

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology, 47*(2), 166–173. https://doi.org/10.1037/h0062855

March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent PsychiatRy, 36*(4), 554–565. https://doi.org/10.1097/00004583-199704000-00019

Marôco, J. (2010a). *Análise de equações estruturais: Fundamentos teóricos, software & aplicações* [Structural Equation Modeling: Theoretical foundations, software and applications]. ReportNumber.

Marôco, J. (2010b). *Análise estatística: Com utilização do SPSS* [Statistical analysis using SPSS] (3rd ed.). Edições Sílabo.

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

Melo, A. (2006). Ansiedade aos exames em contexto universitário [Test Anxiety in an university context] [Master's Thesis]. University of Coimbra.

Mowlem, F. D., Skirrow, C., Reid, P., Maltezos, S., Nijjar, S. K., Merwood, A., Barker, E., Cooper, R., Kuntsi, J., & Asherson, P. (2019). Validation of the mind excessively wandering scale and the relationship of mind wandering to impairment in adult ADHD. *Journal of Attention Disorders, 23*(6), 624–634. https://doi.org/10.1177/1087054716651927

Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M., & Schooler, J. W. (2013). Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00560

Nevitt, J., & Hancock, G. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 353–377. https://doi.org/10.1207/S15328007SEM0803_2

Núñez-Peña, M. I., Suárez-Pellicioni, M., & Bono, R. (2016). Gender differences in test anxiety and their impact on higher education students' academic achievement. *Procedia - Social and Behavioral Sciences, 228*, 154–160. https://doi.org/10.1016/j.sbspro.2016.07.023

Overmeyer, R., & Endrass, T. (2023). Cognitive symptoms link anxiety and depression within a validation of the German State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Clinical Psychology in Europe, 5*(2), Article e9753. https://doi.org/10.32872/cpe.9753

Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2014). When does anxiety help or hinder cognitive test performance? The role of working memory capacity. *British Journal of Psychology, 105*(1), 92–101. https://doi.org/10.1111/BJOP.12009

Pereira, D. S. M., Pereira, A. M., Castanho, T. C., Silva, G. A., Falcão, F., Costa, P., & Pêgo, J. M. (2021). Applicability and validation of the Reaction to Tests Scale (RTT) in a sample of Portuguese medical students. *BMC Psychology, 9*, Article 166. https://doi.org/10.1186/s40359-021-00656-w

Pestana, M. H., & Gageiro, J. N. (2008). *Análise de Dados para Ciências Sociais: A complementaridade do SPSS* [Data analyses for Social Sciences: The complementarity of SPSS] (5th ed). Edições Sílabo.

Ponciano, E., Loureiro, L., Pereira, A., & Spielberger, C. (2005, may). *Características psicométricas e estrutura factorial do TAI de Spielberger em estudantes universitários* [Psychometric characteristics and factorial structure of Spielberger's TAI in university students]. Communication presented at the *Social Action and Counseling in Higher Education – Research and Intervention*, Coimbra, Portugal.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

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

Putwain, D. W., Woods, K. A., & Symes, W. (2009). Personal and situational predictors of test anxiety of students in post-compulsory education. *British Journal of Educational Psychology, 80*(1), 137–160. https://doi.org/10.1348/000709909x466082

Putwain, D. W., Daly, A. L., Chamberlain, S., & Sadreddini, S. (2016). "'Sink or swim': Buoyancy and coping in the cognitive test anxiety–academic performance relationship." *Educational Psychology, 36*(10), 1807–1825. https://doi.org/10.1080/01443410.2015.1066493

Putwain, D. W., Symes, W., & Remedios, R. (2016). The impact of fear appeals on subjective-task value and academic self-efficacy: The role of appraisal. *Learning and Individual Differences, 51*, 307–313. https://doi.org/10.1016/j.lindif.2016.08.042

Putwain, D. W., Symes, W., & Wilkinson, H. M. (2017). Fear appeals, engagement, and examination performance: The role of challenge and threat appraisals. *British Journal of Educational Psychology, 87*(1), 16–31. https://doi.org/10.1111/bjep.12132

Putwain, D. W., Stockinger, K., von der Embse, N. P., Suldo, S. M., & Daumiller, M. (2021). Test anxiety, anxiety disorders, and school-related wellbeing: Manifestations of the same or different constructs? *Journal of School Psychology, 88*, 47–67. https://doi.org/10.1016/J.JSP.2021.08.001

Putwain, D. W., von der Embse, N. P., Rainbird, E. C., & West, G. (2021). The development and validation of a new Multidimensional Test Anxiety Scale (MTAS). *European Journal of Psychological Assessment, 37*(3), 236–246. https://doi.org/10.1027/1015-5759/a000604

Roos, A., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review, 33*(2), 579–618. https://doi.org/10.1007/s10648-020-09543-z

Roos, A., Goetz, T., Krannich, M., Donker, M., Bieleke, M., Caltabiano, A., & Mainhard, T. (2022). Control, anxiety and test performance: Self-reported and physiological indicators of anxiety as mediators. *British Journal of Educational Psychology, 93*(S1), 72–89. https://doi.org/10.1111/bjep.12536

Rosário, P., & Soares, S. (2004). Questionário de Ansiedade face aos Testes (Q.A.T.) [Test Anxiety Questionnaire (T.A.Q.)]. In M. M. Gonçalves, L. Almeida, & C. Machado (Eds.), *Avaliação psicológica: Instrumentos validados para a população portuguesa [Psychological assessment: Validated instruments for the Portuguese population]*. Quarteto.

Rost, D. H., & Schermer, F.J. (1992). "Reactions to Tests" (RTT) and "Manifestations of Test Anxiety" (DAI-MAN): Same or different concepts? In K. A. Hagtvet & T. B. Johnsen (Eds.), Advances in Test Anxiety Research (Vol. 7, pp. 114–129). Swets & Zeitlinger / Swets North America.

Rucklidge, J. J. (2010). Gender differences in Attention-Deficit/Hyperactivity Disorder. *Psychiatric Clinics of North America, 33*(2), 357–373. https://doi.org/10.1016/j.psc.2010.01.006

Salari, N., Ghasemi, H., Abdoli, N., Rahmani, A., Shiri, M. H., Hashemian, A. H., Akbari, H., & Mohammadi, M. (2023). The global

prevalence of ADHD in children and adolescents: A systematic review and meta-analysis. *Italian Journal of Pediatrics, 49*(1), Article 48. https://doi.org/10.1186/s13052-023-01456-1

Salmela-Aro, K. (2011). Stages of adolescence. In B. B. Brown & M. J. Prinstein (Eds.), *Encyclopedia of Adolescence* (Vol. 1, pp. 360–368). Academic Press. https://doi.org/10.1016/B978-0-12-373951-3.00043-0

Salvador, M. C., Matos, A. P., Oliveira, S., March, J. S., Arnarson, E. Ö., Carey, S. C., & Craighead, W. E. (2017). A Escala Multidimensional de Ansiedade para Crianças (MASC): Propriedades psicométricas e análise fatorial confirmatória numa amostra de adolescentes Portugueses [The Multidimensional Anxiety Scale for Children (MASC): Psychometric properties and confirmatory factor analysis in a sample of Portuguese adolescents]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica, 45*(3), 33–46. https://doi.org/10.21865/RIDEP45.3.03

Salvador, M. C. (2009). Ser eu próprio entre os outros: Um novo protocolo de intervenção para adolescentes com fobia social generalizada [To be myself among the others: A new intervention protocol for adolescents with generalized social phobia] [Doctoral dissertation, Faculty of Psychology and Educational Sciences, University of Coimbra, Portugal]. University of Coimbra Digital Repository. https://estudogeral.sib.uc.pt/bitstream/10316/12527/3/Tese%20Doutoramento%20Maria%20do%20C%C3%A9u%20Salvador.pdf

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46*(4), 929–938. https://doi.org/10.1037/0022-3514.46.4.929

Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. (1960). *Anxiety in elementary school children*. John Wiley & Sons.

Sarason, S. B., Hill, K. T., & Zimbardo, P. G. (1964). A longitudinal study of the relation of test anxiety to performance on intelligence and achievement tests. *Monographs of the Society for Research in Child Development, 29*(7), 1–51. https://doi.org/10.2307/1165688

Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *Lancet Child & Adolescent Health, 2*(3), 223–228. https://doi.org/10.1016/S2352-4642(18)30022-1

Segool, N. K., Carlson, J., Goforth, A. N., von der Embse, N., & Barterian, J. (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

Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in Cognitive Sciences, 20*(8), 605–617. https://doi.org/10.1016/j.tics.2016.05.010

Seli, P., Beaty, R. E., Marty-Dugas, J., & Smilek, D. (2019). Depression, anxiety, and stress and the distinction between intentional and unintentional mind wandering. *Psychology of Consciousness: Theory, Research, and Practice, 6*(2), 163–170. https://doi.org/10.1037/cns0000182

Shin, L. M., & Liberzon, I. (2010). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology, 35*(1), 169–191. https://doi.org/10.1038/npp.2009.83

Silva, J., Cunha, M., Galhardo, A., Couto, M., & Massano-Cardoso, I. (2017). Assessment of test anxiety in Portuguese adolescents: Psychometric properties of the Cognitive Test Anxiety Revised (CTAR25). *European Psychiatry, 41*(S1), S434. https://doi.org/10.1016/j.eurpsy.2017.01.424

Smallwood, J., & Schooler, J. W. (2014). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Spielberger, C. D., & Vagg, P. R. (1995). Text anxiety: A transactional process model. In C. D. Spielberger e P. R. Vagg (Eds.), *Text Anxiety: Theory, Assessment, and Treatment* (pp. 3–14). Taylor & Francis.

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Algaze, B., & Anton, W. D. (1978). Examination stress and test anxiety. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and anxiety* (Vol. 5, pp. 167–191). Hemisphere.

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

Taylor, J., & Deane, F. P. (2002). Development of a short form of the Test Anxiety Inventory (TAI). *Journal of General Psychology, 129*(2), 127–136. https://doi.org/10.1080/00221300209603133

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. *American Psychological Association*. https://doi.org/10.1037/10694-000

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test–retest reliability. *PLoS ONE, 8*(9), Article e73990. https://doi.org/10.1371/journal.pone.0073990

Vicente, A. R. F., & Salvador, M. C. (2011). *O Reações aos Testes (RT): Estudos de validação numa amostra de adolescentes portuguesa. [The Reactions to Tests (RT): Validation studies in a Portuguese adolescent sample]*. [Unpublished manuscript]. Faculty of Psychology and Educational Sciences, University of Coimbra, Portugal.

von der Embse, N. P., Kilgus, S. P., Segool, N., & Putwain, D. (2013). Identification and validation of a brief test anxiety screening tool. *International Journal of School & Educational Psychology, 1*(4), 246–258. https://doi.org/10.1080/21683603.2013.826152

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

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

World Health Organization. (2019). *International Classification of Diseases and related health problems* (11th ed.).

Wuthrich, V. M., Belcher, J., Kilby, C., Jagiello, T., & Lowe, C. (2021). Tracking stress, depression, and anxiety across the final year of secondary school: A longitudinal study. *Journal of School Psychology, 88*, 18–30. https://doi.org/10.1016/j.jsp.2021.07.004

Zeidner, M. (2014). Anxiety in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 265–288). Routledge.

Zeidner, M. (1998). *Text anxiety: The State of the Art*. Plenum Press.

Zhao, Y., Summers, R., Gathara, D., & English, M. (2024). Conducting cross-cultural, multi-lingual or multi-country scale development and validation in health care research: A 10-step framework based on a scoping review. *Journal of Global Health*, *14*. https://doi.org/10.7189/jogh.14.04151