von der Embse, N, Putwain, DW and Francis, G

Interpretation and Use of the Multidimensional Test Anxiety Scale (MTAS)

von der Embse, N, Putwain, DW and Francis, G (2021) Interpretation and Use of the Multidimensional Test Anxiety Scale (MTAS). School Psychology, 36 (2). pp. 86-96. ISSN 2578-4218

Interpretation and Use of the Multidimensional Test Anxiety Scale (MTAS)

Nathaniel P. von der Embse, Ph.D.

University of South Florida

David W. Putwain. Ph.D.

Liverpool John Moores University

Gabrielle Francis, M.A.

University of South Florida

This project was funded by a British Academy Award to the second author (SG170738).
The dataset used for this project is publicly available at doi: 10.17632/9dcs5nvwtw.1

Please address correspondence to:
Nathaniel von der Embse, Ph.D.
University of South Florida
College of Education
4202 E. Fowler Avenue
Tampa, FL 33620
natev@usf.edu

Abstract

Test anxiety has proliferated in the era of test-based accountability, however there are limited tools that allow for consistent identification of students at risk. The present study reports on the psychometric evidence and continued development of the Multidimensional Test Anxiety Scale (MTAS). Evidence is presented to support both the interpretation and use of the MTAS. The present sample included over 900 high school aged students across England and Wales that completed a measure of test anxiety and clinical anxiety. Results from confirmatory factor analyses, measurement invariance, and internal consistency support the interpretation of the MTAS. In addition, receiver operator characteristic curve analyses were used to identify initial cut scores to support decision making in applied settings. Results for practice and future research are presented.

Keywords: *test anxiety, assessment, screening, mental health*

Impact and Implications

As schools continue to adopt prevention oriented mental health systems of care, the use of evidence-based assessments tools to inform intervention is paramount. This is particularly true for oft missed problem areas within the internalizing domain such as test anxiety. Results from the present investigation were promising in building necessary evidence for the MTAS to be used within school settings. Further research will continue to inform development efforts and intervention selection.

Introduction

Addressing the internalizing concerns of students has long been a gap in school-based mental health supports (Splett et al., 2019). As schools have become a primary provider for mental health supports (Bruhn et al., 2014), many existing services have focused on externalizing behavior problems that are most disruptive to the classroom environment rather than less noticeable concerns including withdrawal and anxiety (McIntosh et al., 2014). This is particularly problematic, as anxiety is one of the most prevalent mental health problems in children with rates upwards of 40% (Cartwright-Hatton et al., 2006). If left unaddressed, internalizing problems can become more severe with substantial negative influences on long term social and academic outcomes (Suldo et al., 2014). Over 60 years of meta-analytic review has indicated consistent negative relations between a specific type of internalizing concern, test anxiety, and a range of outcomes including academic performance (Hembree, 1988; von der Embse et al., 2018).

In the era of performance-based educational accountability policies, schools are particularly keen to identify areas that may contribute to academic underperformance on high-stakes examinations (von der Embse & Putwain, 2015). Within school settings, there has been an increased focus on "academic anxieties" (Cassady, 2010) which is a class of anxieties that is both *context* and *content* specific. Academic anxieties include a student's anxious response to a type of classroom instruction (i.e., negative motivational instructional tactics or fear appeals; Putwain et al., 2021; Putwain & Best, 2011) or academic content domain (e.g., math anxiety; Hembree, 1990). Test anxiety (TA) an academic anxiety that is characterized by an individual's ongoing emotional, cognitive, and physiological responses to the *perceived* consequences of an evaluative situation (Spielberger & Vagg, 1995). Thus, how a test and the resulting scores are described to students may in turn influence the severity of an anxious response (Putwain & Symes, 2014; von der Embse et al., 2015). These differences are noted in higher levels of test anxiety reported on high-stakes, statewide exams versus typical classroom tests (Segool at al., 2013). However, many schools do not engage in systematic screening for test anxious responses before upcoming exams, and often utilize test anxiety measures in a diagnostic manner and only after a student may be referred for academic underperformance. Without brief and efficient measures, many students may not receive needed universal or targeted intervention supports within an integrated school mental health system (von der Embse et al., 2013).

Approximately 10-20% of students may exhibit significantly high levels of test anxiety (Putwain & Daly, 2014; Thomas et al., 2018). Moreover, students with higher levels of test anxiety consistently indicate lower levels of test performance across exam types and conditions (von der Embse et al., 2019). However, the true prevalence may be unknown as there is a lack of tools with adequate psychometric evidence necessary for widespread use. Modern measurement development should provide evidence to support both a tool's interpretation and use (Kane, 2013) with the ultimate goal of improving outcomes and thus supporting the consequential validity of the assessment method (Messick, 1995). However, many existing test anxiety assessment tools are limited in a number of important ways. Specifically, they are not reflective of contemporary theoretical understanding of test anxiety, lack data to support equivalence of measurement across demographic groups, and were developed for research purposes (i.e., supporting interpretation) without evidence to support use in applied decision making.

Theoretical advancements are necessary to drive modern assessment development. That is, tools should reflect improvements in understanding phenomena over time (Rajagopalan & Gordon, 2016). Early theories of test anxiety were largely unidimensional (Mandler & Sarason,

1952) and later included cognitive and affective components (Liebert & Morris, 1967). Subsequent advancements introduced cognitive interference (Friedman & Bendas-Jacob, 1997), motivational (Pekrun at al.,, 2004), and social components (Lowe et al., 2008). Following the widespread adoption of test-based accountability policies, new theoretical developments have included the appraisal of an evaluation as a threat such as the perceived use of test performance to inform important life outcomes (e.g., admission to University; Segool et al., 2014; Ziedner & Matthews, 2005). Cassady (2010) expanded upon the contextual importance of test anxiety measurement by differentiating anxious responses before, during, and after the exam. Unfortunately, the development of test anxiety measures has not advanced in concert with more recent progressions in theory. For example, the most widely used assessment (*Test Anxiety Inventory,* Spielberger, 1980), is now more than 40 years old. Other tools do not consider student perceptions of test use (e.g., *Children's Test Anxiety Scale*; Wren & Benson, 2004) despite the clear difference in the manifestation of anxiety depending on the instructional context or how a test is used (Segool et al., 2013).

Although important, reflecting contemporary theory is not sufficient to support interpretation and use of an assessment. Increasing attention has been paid to the importance of test scores that reflect the same construct and meaning for all test takers within the intended population (AERA, APA, NCME, 2014). Moreover, the success of evidence-based interventions in a school setting is dependent on data allowing for valid inferences (Kratochwill & Stoiber, 2002). Evaluating measurement invariance (or equivalence) thus is necessary to support use within diverse populations, including with specific demographic characteristics (e.g., gender, socio-economic status; Pendergast et al., 2017). This is particularly important given the base rate differences within test anxiety as reported in meta-analytic work (see Hembree, 1988, von der

Embse et al., 2018). Although several TA tools have provided preliminary evidence for similar factor structures in different populations (e.g., US and Singapore, *Test Anxiety Scale for Elementary Students*; Lowe et al., 2011), there is a dearth of psychometric evidence that has evaluated the multiple types of invariance (e.g., Pendergast et al., 2017) to support use within various populations. Other limitations from existing tools include reporting of internal consistency without consideration of the hierarchical (or nested) structure of the data (see Gignac & Watkins, 2013). These data would be necessary to provide evidence for the interpretation of subscales as unique (variance) and apart from the total scale.

Lastly, there is a critical need for empirical research to inform the consequential validity of assessments (Messick, 1995). This is particularly true for TA measures. For example, the *Friedben Test Anxiety Scale* (FTAS; Friedman & Bendas-Jacob, 1997) and a screening version of the FTAS (von der Embse et al., 2013) provide preliminary evidence for interpretation (via factor analyses) yet no evidence (i.e., cut scores) to support use within decision-making. The *Cognitive Test Anxiety Scale* (CTAS; Cassady & Johnson, 2002) is a notable exception with the provision of cut-scores developed via latent class analyses (Thomas et al., 2018). However, additional research is necessary to determine how these tools inform treatment outcomes with school-aged populations. Consistent with the interpretation and use argument posited by Kane (2013), psychometric research should 1) corroborate factor structures in new populations and in consideration of discriminant and convergent variables, 2) evaluate score performance and interpretation across diverse populations, and 3) provide evidence to support decisional inferences.

In recognition of the limitations of existing measures, the *Multidimensional Test Anxiety Scale* (MTAS; Authors, 2020) was developed over the course of three years and inclusive of nearly 9,000 students. The MTAS was developed to reflect contemporary theory and the *context* of TA with considerations of perceived test importance. Through a rigorous content development process, the MTAS sought to reconcile construct confusion across TA measures where the same term was being used to describe different constructs, and different terms were used to reflect the same constructs (Kelley, 1927). In addition, a specific goal of the MTAS development was to produce a brief and efficient tool that could be used for a variety of purposes (including screening) and inform school decision-making. Preliminary evidence across two studies has supported factor structure, internal consistency, and test-retest reliability in multiple samples, in addition to discriminant and convergent validity (Authors, 2020). The next phase in development is to support the inferences drawn from MTAS scores, including the evaluation of measurement invariance and creation of cut scores to inform decision-making.

Current Study

To realize the potential of a preventative and integrated school mental health system, evidence-based decision-making tools are necessary. This is particularly true for measures of internalizing problems, including TA, that are not developed and validated with the same frequency as measures of externalizing concerns. The primary aims of the current investigation are threefold including 1) an evaluation of factor structure and reliability after adaption of two items, and in relation to negative (clinical anxiety symptoms) outcomes, 2) the measurement invariance across demographic groups, and 3) the generation of initial cut scores to inform how the scale may be used in applied settings.

Method

Participants and Procedure

The sample consisted of 918 participants (male = 217, female = 694, chose not to disclose = 7) with a mean age of 15.76 years (*SD* = 1.13) drawn from eight secondary schools located in England and Wales. Two of the schools were single gender girls' schools (accounting for sample skew towards females) and the remainder were coeducational. Students were in either the final two years of compulsory secondary education (Year 10, *n* = 100; Year 11, *n* = 481) or the two years of post-compulsory upper secondary education (Year 12, *n* = 158; Year 13, *n* = 179). Participants reported their ethnic background as British Asian (*n* = 29), British Black (*n* = 48), British White (*n* = 802), dual heritage (*n* = 21), and other (*n* = 18). One hundred and forty participants (15.3%) were eligible for free school meals, a proxy for low income.

Data were collected in schools during a period of the school timetable used for administrative purposes, so as not to interfere with routinely scheduled lessons, and took approximately ten minutes to complete. Permission to undertake the study was provided by an institutional research ethics committee and written institutional consent was provided by the Head or Principal of each school. Data were collected electronically using an online survey tool that prompted participants if they had missed an answer; accordingly, there were no missing data. Individual participants were provided with an information sheet explaining that individual consent was implied by completing the online questionnaire. Withdrawal of participation during data collection was achieved by closing the web browser. In order to offer retrospective withdrawal, participants were asked to provide a memorable word on the survey that could be passed to the research team via the school liaison to delete responses.

Measures

Test Anxiety

Test anxiety was measured using the 16-item *Multidimensional Test Anxiety Scale* (MTAS; Authors, 2020). This instrument consists of two cognitive subscales, namely worry (e.g., ‘Before a test/ exam, I am worried I will fail’) and cognitive interference (e.g., ‘During tests/exams, I forget things that I have learnt’), and two affective-physiological items, namely tension (e.g., Even when I have prepared for a test/ exam I feel nervous about it’) and physiological indicators (e.g., ‘My heart races when I take a test/exam’). Each subscale comprises four items. Participants responded to items on a five-point scale (1 = ‘strongly disagree, 3 = ‘neither’, 5 = ‘strongly agree’); a higher score represents greater TA. MTAS scores can be represented as a single total TA score or as four subscale scores, depending on substantive research questions. MTAS scores have shown strong internal consistency (ωs = .85 - .91), factorial validity (items loading on target factors λs = .46 to .92), and predictive validity where higher TA is related to elevated mental health risk (*r*s = .13 to .46), and lower examination performance (*r*s = .01 to .41), and wellbeing (*r*s = -.03 to -.33) (Authors, 2020).

In the present study, we replaced two MTAS items to reduce the need for correlated residual variances as identified in the original study. Two of the original MTAS physiological indicators items (items 4 and 16), and two the original MTAS worry items (items 5 and 13) had overlapping wording necessitating correlated residual variances. One worry, and one physiological indicators, item were replaced with alternatives that did not show any wording overlap from the larger item pool used by Authors et al. (2020) in an exploratory factor analysis (EFA). Item 16, ‘My hand shakes while I am taking a test/exam’, was replaced with ‘During a test/ exam, my muscles are tight’. Item 5, ‘I am afraid of writing the wrong answer during a test/exam’, was replaced with ‘During tests/ exams, I worry about the consequences of failing’). We also modified the temporal referent of worry item 13 from ‘After a test/ exam…’ to ‘During tests/ exams…' so that two of the four worry items denoted worry during a test or an exam (of the remaining two items, one referred to worry *before*, and the other worry *after*, a test/ exam). All MTAS items used in the present study are reproduced in the Appendix.

Clinical Anxiety

Clinical anxiety was measured using the six-item Generalized Anxiety (e.g., 'I worry that something bad will happen to me') and the nine-item Panic ('I suddenly become dizzy or faint when there is no reason for this') subscales from the *Revised Children's Anxiety and Depression Scale* (RCADS: Chorpita et al., 2005). Participants responded in a 4-point scale (0 = 'Never', 1 = 'Sometimes', 2 = 'Often', and 3 = 'Always') meaning a higher score indicates greater anxiety. The RCADS is a widely used instrument (Piqueras et al., 2017) and subscales have shown excellent psychometric properties including factorial (items loading on target factors λs = .69 to .89 for Generalized Anxiety, and λs = .52 to .86 for Panic), and convergent (*r*s = .52 to .74 for Generalized Anxiety, and .69 to .78 for Panic, with the anxiety subscale of the Depression, Anxiety and Stress Scale Short Version), validity, and internal consistency (αs = .84 to .88 for GAD and .74 to .86 for Panic) (Donnelly et al., 2019).

Analytic Strategy

The analytic strategy followed three stages. First, was to check the descriptive and psychometric properties of constructs using confirmatory factor analyses (CFAs) to check for model fit and generate latent bivariate correlations, reliability analyses (using McDonald's ω where $\omega < .7$ is considered acceptable; Dunn et al., 2014). In addition, to examine the contribution of first and second-order factors to true score variance, omega-hierarchical and its subscales were computed as well as related *H* indexes (Reise et al., 2013). This was particularly germane to the MTAS where we had modified two items to reduce the reliance on correlated residual variance. Second, was to check the invariance of the MTAS for gender and free school meals, as a proxy for low income, in a series of multi-group CFAs. The model fit of CFAs was guided by using the following indices: Root Mean Square Error of Approximation (RMSEA), Standardized Root Means Square Residual (SRMR), Comparative Fit Index (CFI), and the Tucker-Lewis index (TLI). A RMSEA $\leq .08$, SRMR $\leq .06$, and CFI/ TLI indices $\geq .95$, has been interpreted as showing a good fit to the data (Hu & Bentler, 1999). Some analysts have cautioned, however, that such values may be overly stringent for complex naturalistic data (Heene, et al., 2011) and CFI/ TLI indices $\geq .90$ are indicative of an 'approximate' fit (McDonald & Marsh, 1990). Changes in model fit criteria for the invariance testing adhered to recommended best practices (see Pendergast et al., 2017). Third, was to establish cut-scores for MTAS for generalized anxiety and panic using Receiver-Operator-Characteristic (ROC) curve analysis.

Results

Descriptive Statistics and Bivariate Correlations

Descriptive statistics as well as estimates of internal consistency are reported in Table 1. Although individual variables were normally distributed (skewness and kurotsis within ±1) multivariate normality could not be assumed (Mardia's multivariate skewness $b = 77.08$, $p < .001$, and kurtosis $b = 1151.28$, $p < .001$[1]). All scales demonstrated excellent internal consistency across McDonald's omega (McDonald's ωs $\geq .83$) as well as coefficient *H* indices (Reise et al., 2013). The proportion of variance attributable to schools, the intraclass correlation coefficient, or ICC1 statistic (ρ_I), was negligible. CFAs for constructs are reported in Table 2. Test anxiety was specified as a higher-order model comprising of lower lower-order factors (worry, cognitive interference, tension, and physiological indicators) and one higher-order factor (test anxiety). Generalized anxiety, and panic, were specified as univariate scales. To account for the deviation in multivariate normality all models were estimated using maximum likelihood with robust standard error (MLR) in M*plus* v. 8.3 (Muthén & Muthén, 2017). For the two RCADS subscales, a residual correlation was added to a pair of Generalized Anxiety items with similar wording referring to 'bad things' (items 2 and 5) and residual correlations were added to pairs of Panic items with similar wording for 'feeling scared' (items 5 and 9) and 'heart beating fast' (items 2 and 8). Model fit indices are reported in Table 2. All models showed a good fit to the data (with the caveat that although for panic RMSEA = 085, the other model fit indices suggested a good fit).

In order to generate latent bivariate correlations a CFA was conducted with all constructs included in a single measurement model that also included gender (0 = male, 1 = female; students that did not specify gender were treated as missing data and handled using full information maximum likelihood), age, and free school meals (FSM; 0 = no FSM, 1 = FSM) As manifest constructs (see Figure 1). The CFA was estimated using MLR and showed a reasonable fit to the data. The factor loadings, however, were strong (λs >.5; see Table 3) and there were no obvious signs of model misspecification (e.g., low variance explain by items). Guided by modification indices, the plausible reason for this model showing an acceptable, rather than good, fit to the data is low-level cross-loading to non-target factors for anxiety TA, generalized anxiety, and panic items, artificially constrained to zero in the CFA (see Marsh et al., 2020). Accordingly, estimated model parameters were treated as robust and we proceeded to interpret the model; it is likely this is an example of model fit indices being influenced by 'messy' naturalistic data that does not neatly load onto a single factor (see Heene et al., 2011).

Latent bivariate correlations are reported in Table 4. TA correlated positively with both generalized anxiety and panic subscales on the RCADS. Female students reported greater test anxiety, generalized anxiety, and panic, and students eligible for FSM reported greater generalized anxiety. We do not attach any interpretative significance to the positive correlations between TA and age, and between FSM, TA and panic (*r*s <.1 are considered negligible; Cohen, 1988).

Multi-Group Tests for Gender and Free School Meal Invariance

The purpose of this analysis is to establish whether the measurement properties of the MTAS were equivalent for key groups of participants, namely males and females, and those in receipt of FSM or not (indicative of family circumstances characterized by greater or lesser deprivation respectively). Four models were tested sequentially. First, a configural model, where factor loadings, item intercepts, and item residual variances, are allowed to vary between the groups in question. Second, a metric invariance model to constrain factor loadings for the two groups as equivalent. Third, a scalar invariance model to constrain item intercepts between the two groups as equivalent (strong invariance). Fourth, a residual invariance model, to constrain item residual variance between the two groups as equivalent, (strict invariance). A decline in model fit of >.015 for the RMSEA, and >.01 for the CFI/ TLI, indicates non-invariance (Chen, 2007; Cheung & Rensvold, 2002).

All models were estimated using MLR in M*plus* v. 8.3 (Muthén & Muthén, 2017). Invariance tests are reported in Table 5. Configural models for gender and FSM showed a good fit to the data and there was no substantive decline in model fit when successive constraints were applied to factor loadings, item intercepts, and item residual variances. The MTAS showed strict invariance for gender and FSM. There were large differences in the latent factor mean of TA scores for female participants of 0.94, p <.001 (Male M = 48.17, SD = 12.49; Female M = 57.64. SD = 11.22; d = .80) and small differences for participants in receipt of FSM of 0.28, p =.007 (No FSM M = 54.83, SD = 12.02; FSM M = 58.48. SD = 12.88; d = .29).

Cut Score Development

Receiver operating characteristic curve (ROCC) analyses (see Figure 2) were employed to generate cut scores and evaluate the resulting diagnostic accuracy. Cut scores were evaluated for accuracy in relation to clinical outcomes of internalizing concerns (e.g., anxiety, panic) on the RCADS. The area under the curve (AUC) was calculated for the MTAS total scale and subscales. The AUCs were considered between 0.50 and 0.70 as reflective of low accuracy, 0.70–0.90 moderate, and 0.90–1.00 high (Streiner & Cairney, 2007). A 95% confidence interval (CIs) was used for all AUC calculations. The diagnostic accuracy was then calculated with four metrics including sensitivity (SE; proportion of individuals who were correctly identified with high TA), specificity (SP; proportion of individuals who were correctly identified without high TA) as well as positive and negative predictive validity. Although no gold standard exists, SE values of ≥ 0.80 were considered acceptable and values ≥ 0.70 borderline, and SP values of ≥ 0.70 and ≥ 0.60 were considered acceptable and borderline, respectively (Kilgus et al., 2014). For each SE and SP value, 95% CIs were calculated with 2000 stratified bootstrapped replicates.

As previously noted, cut scores were selected and prioritized to afford optimal sensitivity and then specificity when possible. Cut scores were derived from the MTAS Total and four subscales. Overall, AUC values consistently fell within the moderate range across all the subscales. The identified cut scores, as well as AUC and sensitivity and specificity are reported within Table 5. The identified cut scores of 58 and 60 on the MTAS Total scale is on the 72nd and 75th percentile respectively. With a cut score of 60, approximately 35% of the current study sample would be considered highly test anxious. The MTAS total score consistently reflected the highest AUC as well as sensitivity and specificity across scales. All of the subscale cut scores yielded borderline to acceptable sensitivity as predictors of anxiety. The Tension subscale cut score yielded unacceptable specificity in relation to anxiety while the rest of the subscales cut scores were borderline or acceptable. All of the MTAS subscale cut scores indicated borderline to acceptable sensitivity as predictors of Panic. Worry and Tension cut scores were below borderline specificity as predictors of Panic.

Discussion

As schools increasingly adopt prevention-oriented systems of mental health intervention, there is a critical need to develop evidence-based assessment tools particularly within the internalizing domain. TA is one such internalizing domain with limited tools that support use within applied settings. The primary goal of the present investigation was to advance the psychometric evidence for both interpretation and use of the Multidimensional Test Anxiety Scale. Results from the current study were promising towards this goal.

An important developmental process to support the interpretation of an assessment involves the confirmation of factor structures in new samples. To that end, confirmatory factor analyses supported the factor structure of the MTAS within a large new sample and after revision to two test items. In addition, reliability analyses further supported the MTAS model of interpretation with strong internal consistency of MTAS subscales. These results are encouraging yet should be interpreted with caution. For example, although the factor structure was supported, additional research will be necessary to confirm this structure with the same participants across time as well as new participants across ages and educational settings.

Although differences were noted in mean scores across genders, strict measurement invariance was reported and suggests the MTAS to perform with equivalence across genders and socio-economic status. However, additional analyses in future studies will be necessary to examine potential predictive bias across demographic groups (including age and race) and to determine the presence of measurement equivalence in new settings. This will be particularly important given the cultural variation in the expression of internalizing concerns (Crijnen et al., 1997; Kim et al., 2019). In addition, multi-level or multi-group methods of measurement invariance with large samples may inform future research and account for nesting within classrooms and schools (Pendergast et al., 2017).

Lastly, an initial set of cut scores were generated in relation to clinically relevant internalizing behaviors (anxiety, panic) with acceptable results for the total score on the MTAS and several subscales. Cut scores slightly varied depending on the clinical outcome of interest. Across outcomes, the MTAS total score performed defensibly with regards to sensitivity and specificity. These cut scores may be considered for exploratory use with students self-reporting risk status for TA. However, and as aforementioned, research will be necessary to confirm these cut scores within new populations, and with similar outcome measures (e.g., the *Test Anxiety Inventory*). Results from the present investigation indicate that between 35-42% of students were considered to be highly test anxious, which is higher than previous estimates (Putwain & Daly, 2014) yet may be reflective of an overall trend towards higher levels of test anxiety since the widespread adoption of high-stakes examinations (von der Embse et al., 2017).

Taken together, this study represents an important next step in the development of the MTAS. The abbreviated nature of the MTAS (16 items) lends its use within schoolwide screening efforts whereas other test anxiety scales were either too lengthy (e.g., >40 items, *Cognitive Test Anxiety Scale;* Cassady, 2002*)* or had evidence for interpretation but no defensible cut scores to inform use (e.g., *Brief-FTAS*; von der Embse et al., 2013). Moreover, the MTAS incorporates contemporary theory that is reflective of the multiple dimensions of TA, while including items that reflect construct manifestation across the examination cycle (before, during, after; Cassady, 2010). Additional research will be needed to continue to corroborate MTAS interpretation and use evidence, and ultimately inform best practices to inform school-based intervention.

Limitations

There are multiple limitations to the present investigation. First, although relatively large, the sample was collected from schools in England and Wales. Previous research has demonstrated notable differences in how an educational setting may provoke test anxious responses (see Segool et al., 2013) Therefore, additional research will be necessary to determine the potential influence on environmental variables (e.g., high-stakes examinations across Eastern and Western cultures) on the manifestation of TA on the MTAS. Relatedly, future investigations should continue to examine measurement equivalence with new demographic groups. Second, there is a lack of evidence to support the integrity of the assessment process. That is, how an assessment is described may influence the subsequent responses, particularly for measures of internalizing behaviors (von der Embse et al., 2015). Research should evaluate the influence of assessment training that may ensure consistent levels of construct knowledge (Evans et al., 2006). Lastly, psychometric evidence will be essential to confirm the stability of decisions made when using the MTAS. This includes determining if the MTAS identifies similar students at different times within the academic year, or if the students identified during the Fall remain the same in subsequent screenings (assuming no intervention). For example, latent profile analyses may be informative to identifying profiles of respondents and the subsequent stability across time (Herman et al., 2018).

Endnote

[1] Tests of multivariate normality were estimated in M*plus* using TECH13 which is based on the definition provided in Mardia et al. (1979).

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Authors, 2020.

Bruhn, A.L., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional and behavioral screening practices in K–12 schools. *Education and Treatment of Children 37*(4), 611-634. doi:10.1353/etc.2014.0039.

Cartwright-Hatton, S., McNicol, K., & Doubleday, E. (2006). Anxiety in a neglected population: Prevalence of anxiety disorders in pre-adolescent children. *Clinical Psychology Review, 26*(7), 817-833. https://doi.org/10.1016/j.cpr.2005.12.002.

Cassady, J.C. (2010). *Anxiety in schools: The causes, consequences, and solutions for academic anxieties*, Vol. 2 Peter Lang.

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

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: a multimethod approach. *Journal of Personality and Social Psychology, 80*, 814-833. https://doi.org/10.1037/0022-3514.80.5.814

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chorpita, B. F., Moffitt, C., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy, 43*, 309-322. https://doi.org/10.1016/j.brat.2004.02.004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed).* Erlbaum.

Crijnen, A. A., Achenbach, T. M., & Verhulst, F. C. (1997). Comparisons of problems reported by parents of children in 12 cultures: total problems, externalizing, and internalizing. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(9), 1269-1277.

Donnelly, A., Fitzgerald, A., Shevlin, M., & Dooley, B. (2018). Investigating the psychometric properties of the Revised Child Anxiety and Depression Scale (RCADS) in a non-clinical sample of Irish adolescents. *Journal of Mental Health, 15*, 1–12. https://doi.org/10.1080/09638237.2018.1437604 Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399-412.doi: 10.1111/bjop.12046

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

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: a cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, *16,* 319-336. doi: 10.1037/a0024917

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

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46. doi:10.2307/749455

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6,* 1–55. doi: 10.1080/10705519909540118

Kane, M.T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, *50*(1), 1-73. https://doi.org.ezproxy.lib.usf.edu/10.1111/jedm.12000

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: World Book Company

Kilgus, S. P., Riley-Tillman, T. C., Chafouleas, S. M., Christ, T. J., & Welsh, M. E. (2014). Direct behavior rating as a school-based behavior universal screener: Replication across sites. *Journal of School Psychology, 52*(1), 63-82. https://doi.org/10.1016/j.jsp.2013.11.002

Kim, J. H., Tsai, W., Kodish, T., Trung, L. T., Lau, A. S., & Weiss, B. (2019). Cultural variation in temporal associations among somatic complaints, anxiety, and depressive symptoms in adolescence. *Journal of Psychosomatic Research*, *124*, 109763.

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of Division 16 and the society for the study of school psychology task force. *School Psychology Quarterly, 17*(4), 341. https://doi.org/10.1521/scpq.17.4.341.20872

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

Loderer, K., Vogl, E., & Pekrun, R. (2016, August 24-27). Students' well-being at school revisited: Development and initial validation of a unidimensional self-report scale [Conference session]. International Conference on Motivation (ICM), Thessaloniki, Greece. http://www.psy.auth.gr/el/conference/programme.

Lowe, P. A., Ang, R. P., & Loke, S. W. (2011). Psychometric analyses of the Test Anxiety Scale for Elementary Students (TAS-E) scores among Singapore primary school students. *Journal of Psychopathology and Behavioral Assessment, 33*(4), 547.

Lowe, P. A., Lee, S. W., Witteborg, K. M., Prichard, K. W., Luhr, M. E., Cullinan, C. M., & 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*, 215–230. doi:10.1177/0734282907303760

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

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1979). *Multivariate analysis*. Academic Press.

Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (in press). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM) & set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research, 55,* 102-119. https://doi.org/10.1080/00273171.2019.1602503

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107*, 247–255. https://doi.org/10.1037/0033-2909.107.2.247

McIntosh, K., Ty, S. V., & Miller, L. D. (2014). Effects of school-wide positive behavioral interventions and supports on internalizing problems: Current evidence and future directions. *Journal of Positive Behavior Interventions, 16*(4), 209-218.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational measurement: Issues and practice, 14*(4), 5-8. https://doi.org/10.1111/j.1745-3992.1995.tb00881.x

Muthén, L. K., & Muthén, B. O. (2017). M*plus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the Test Emotions Questionnaire (TEQ). *Anxiety, Stress & Coping, 17*(3), 287-316. https://doi.org/10.1080/10615800412331303847

Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology, 60*, 65-82. https://doi.org/10.1016/j.jsp.2016.11.002

Piqueras, J.A., Martín-Vivar, M., Sandin, B., San Luis, C., Pineada, D., (2017). The Revised Child Anxiety and Depression Scale: a systematic review and reliabilitygeneralization meta-analysis. *Journal of Affective Disorders, 218*, 153–169. https://doi.org/10.1016/j.jad.2017.04.022.

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

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

Putwain, D. W., Symes, W. (2014) The perceived value of maths and academic self-efficacy in the appraisal of fear appeals used prior to a high-stakes test as threatening or challenging. *Social Psychology of Education* 17, 229–248. doi: 10.1007/s11218-014-9249-7

Putwain, D.W., Symes, W., Nicholson, L.J., & Remedios, R. (2021). Teacher motivational messages prior to examinations: What are they, how are they evaluated, and what are their educational outcomes? In A.J. Elliot (Ed.), *Advances in motivational science*. Academic Press.

Rajagopalan, K., & Gordon, E. W. (2016). *The testing and learning revolution: The future of assessment in education*. Springer.

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

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

Spielberger, C. D. (1980). Preliminary professional manual for the Test Anxiety Inventory (TAI). Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R. and Jacobs, G.A. (1983). Manual for the *State-Trait Anxiety Inventory* (Form Y). Palo Alto, CA: Mind Garden.

Spielberger, C. D., & Vagg, P. R. (1995). Test anxiety: A transactional process model. In C. D. Spielberger & P. R. Vagg (Eds.), *Series in clinical and community psychology. Test anxiety: Theory, assessment, and treatment* (p. 3–14).

Splett, J. W., Garzona, M., Gibson, N., Wojtalewicz, D., Raborn, A., & Reinke, W. M. (2019). Teacher recognition, concern, and referral of children's internalizing and externalizing behavior problems. *School Mental Health*, *11*(2), 228-239.

Suldo, S. M., Gormley, M. J., DuPaul, G. J., & Anderson-Butcher, D. (2014). The impact of school mental health on student and school-level academic outcomes: Current status of the research and future directions. *School Mental Health, 6*(2), 84-98. doi:10.1007/s12013-9116-2

Thomas, C. L., Cassady, J. C., & Finch, W. H. (2018). Identifying severity standards on the cognitive test anxiety scale: Cut score determination using latent class and cluster analysis. *Journal of Psychoeducational Assessment, 36*(5), 492-508. http://dx.doi.org/10.1177/0734282916686004.

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

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. P., & Putwain, D. W. (2015). Examining the context of instruction to facilitate student success. *School Psychology International*, *36*(6), 552-558. https://doi.org/10.1177/0143034315612144

von der Embse, N. P., Schultz, B. K., & Draughn, J. D. (2015). Readying students to test: The influence of fear and efficacy appeals on anxiety and test performance. *School Psychology International*, *36*(6), 620-637. https://doi.org/10.1177/0143034315609094

Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress, and Coping*, *17*(3), 227–240. doi:10.1080/10615800412331292606

Zeidner, M., Matthews, G. (2005). Evaluation Anxiety. In: Elliot, A.J., Dweck, C.S. (Eds.), *Handbook of Competence And Motivation* (pp. 141–163). Guilford Press, London.

Table 1

Descriptive Statistics for Test Anxiety, Generalized Anxiety, and Panic

	Scale Range	Mean	SD	ω [95% CIs]	*H*	ρ_I	Skewness	Kurt
Test Anxiety	16 - 80	55.45	12.21	.75 [.75, .75][a]	.97	.01	-.54	.20
Worry	4 – 20	15.12	3.43	.84 [.82, .85][b]	.90	.02	-.84	.58
Cognitive Interference	4 – 20	14.06	3.43	.83 [.82, .85][b]	.89	.03	-.53	-.03
Tension	4 – 20	15.13	3.62	.88 [.87, .90][b]	.91	.01	-.95	.63
Physiological Indicators	4 - 20	11.08	4.06	.84 [.82, .85][b]	.88	<.01	.12	-.73
Generalized Anxiety	0 - 18	6.26	3.84	.91 [.91, .92]	.94	.01	.50	-.61
Panic	0 - 27	8.64	6.87	.93 [.92, .93]	.92	<.01	.85	-.14

Note. [a] Omega Hierarchical (ωH). [b] Omega Hierarchical Subscale (ωHS).

Table 2
Confirmatory Factor Analyses for Test Anxiety, Generalized Anxiety, and Panic

	χ^2 (df)	RMSEA	SRMR	CFI	TLI
Test Anxiety	376.04 (99)	.055	.050	.959	.950
Generalized Anxiety	35.99 (8)	.062	.023	.986	.974
Panic	190.69 (25)	.085	.033	.956	.937
Combined Measurement Model	2051.35 (511)	.058	.066	.909	.901

Note. χ^2 of all models statistically significant at p <.001. df = degrees of freedom, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Means Square Residual, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index.

Table 3
Standardized Factor Loadings from the Combined Measurement Model

Item		TA	W	CI	T	PI	GA	P
1.	Before a test/ exam, I am worried I will fail.		.78					
2	I forget previously known material before taking a test/exam.		.74					
3.	Even when I have prepared for a test/ exam I feel nervous about it.		.83					
4.	Before I take a test/ exam my hand trembles.		.68					
5.	During tests/ exams, I worry about the consequences of failing.			.72				
6.	I forget facts I have learnt during tests/exams.			.86				
7.	I feel tense before taking a test/exam.			.86				
8.	My heart races when I take a test/exam.			.51				
9.	After a test/exam, I am worried I have failed.				.75			
10.	During tests/exams, I forget things that I have learnt.				.79			
11.	Just before I take a test/exam, I feel panicky.				.85			
12.	During a test/ exam I experience stomach discomfort.				.82			
13.	During a test/ exam, I worry that I gave the wrong answers.					.75		
14.	During tests/exams, I find it hard to concentrate.					.82		
15.	Before a test/exam, I feel nervous.					.72		
16.	During a test/ exam, my muscles are tight					.73		
W		.92						
CI		.59						
T		.93						
PI		.90						
1.	I worry about things						..60	
2.	I worry that bad things will happen to me						.66	
3.	I worry that something awful will happen to someone in my family						.90	
4.	I worry that something bad will happen to me						.91	
5.	I worry about what is going to happen						.74	
6.	I think about death						.60	
1.	When I have a problem I get a funny feeling in my stomach							.60
2.	When I have a problem, my heart beats really fast							.72
3.	I suddenly start to tremble or shake when there is no reason for this							.83
4.	When I have a problem, I feel shaky							.76
5.	All of a sudden I feel scared for no reason at all							.81

6.	I suddenly become dizzy or faint when there is no reason for this	.66
7.	I suddenly feel as if I can't breathe when there is no reason for this	.82
8.	My heart suddenly starts to beat too quickly for no reason	.83
9.	I worry that I will suddenly get a scared feeling when there is nothing to be afraid of	.81

Note. TA = Test Anxiety, W = Worry, CI = Cognitive Interference, T = Tension, PI = Physiological Indicators, Generalized Anxiety, and P = Panic

Table 4
Latent Bivariate Correlations for Test Anxiety, Generalized Anxiety, Panic, Gender, Age, and FSM

	1.	2.	3.	4.	5.	6.
1. Test Anxiety	—	.67***	.62***	.37***	.08*	.09**
2. Generalized Anxiety		—	.81***	.25***	-.01	.15***
3. Panic			—	.25***	.01	.09**
4. Gender				—	—	—
5. Age					—	—
6. FSM						—

Note. FSM = Free School Meals.

Table 5
Gender and FSM Invariance Tests for the Multidimensional Test Anxiety Scale

	χ^2 (df)	RMSEA	SRMR	CFI	TLI	Δ RMSEA	ΔCFI	ΔTLI
Gender Invariance:								
Configural	504.78 (200)	.058	.054	.952	.942	—	—	—
Metric Invariance	553.27 (215)	.059	.068	.947	.940	+.001	-.005	-.002
Scalar Invariance	585.06 (227)	.059	.071	.943	.940	.000	-.004	.000
Residual Invariance	630.73 (243)	.059	.079	.939	.940	.000	-.004	.000
FSM Invariance:								
Configural	497.56 (200)	.057	.053	.957	.948	—	—	—
Metric Invariance	514.32 (215)	.055	.057	.957	.952	+.002	.000	+.004
Scalar Invariance	526.95(227)	.054	.057	.957	.954	+.001	.000	+.002
Residual Invariance	542.81(243)	.052	.060	.957	.957	+.002	.000	+.003

Note. χ^2 of all models statistically significant at p <.001. FSM = free school meals. RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Means Square Residual, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, Δ = change.

Table 6

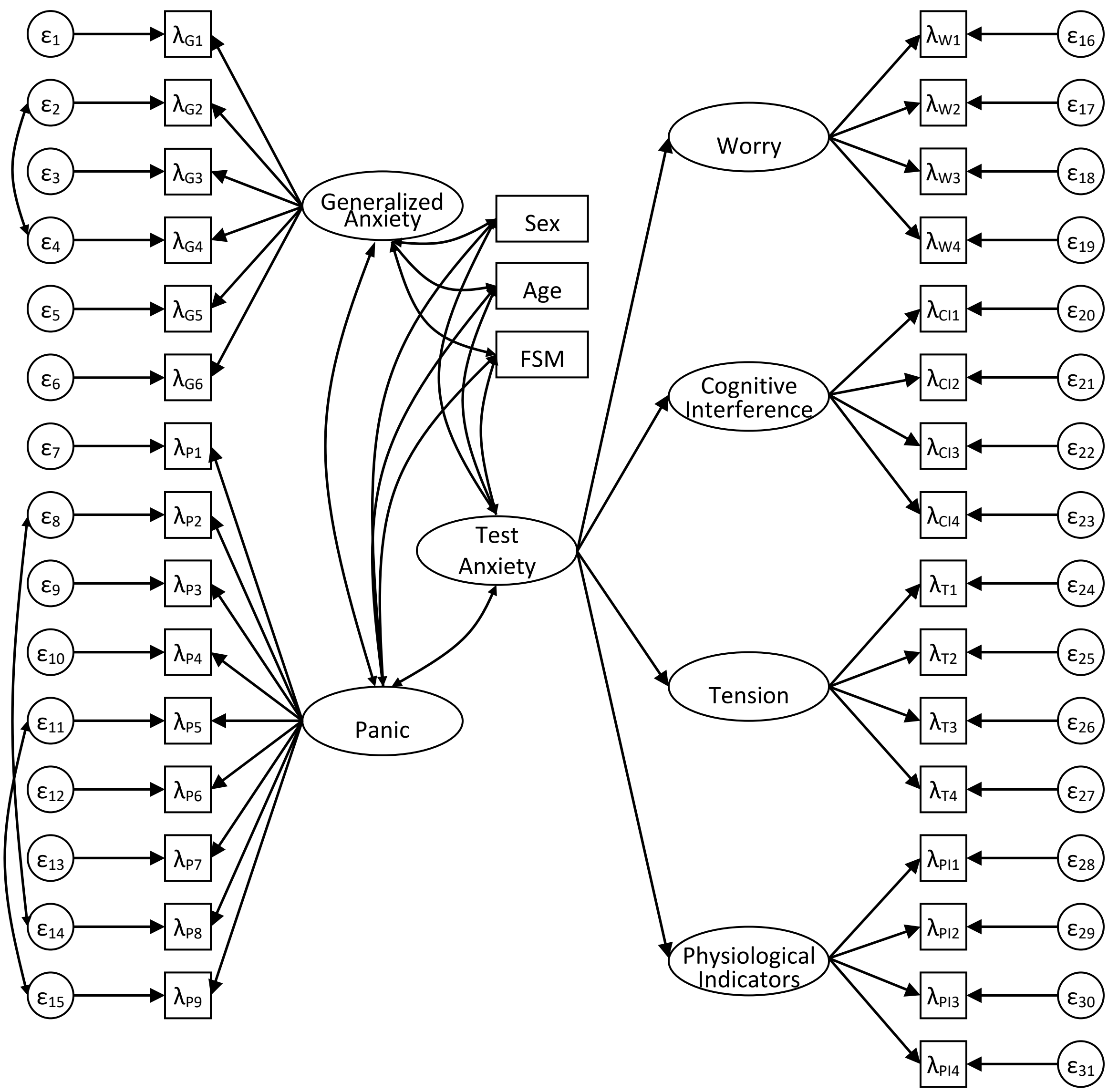
Conditional probability statistics indicative of the diagnostic accuracy of Multidimensional Test Anxiety Scale domain cut scores relative to GAD and Panic Scales

Outcome	Predictor	AUC (95% CI)	Cut Score	SE (95% CI)	SP (95% CI)	PPV	NPV
Anxiety							
	MTAS Total	0.82	58	0.79	0.72	.62	.86
	Worry	0.78	16	0.82	0.62	.77	.89
	Cognitive Interference	0.71	15	0.71	0.63	.52	.79
	Tension	0.77	16	0.82	0.58	.53	.85
	Physiological Indicators	0.78	12	0.71	0.68	.56	.80
Panic							
	MTAS Total	0.84	60	0.80	0.74	.52	.91
	Worry	0.76	16	0.84	0.56	.40	.91
	Cognitive Interference	0.73	15	0.74	0.60	.39	.87
	Tension	0.77	16	0.85	0.53	.39	.91
	Physiological Indicators	0.84	13	0.79	0.76	.54	.91

Note. AUC = area under the curve, SE = sensitivity, SP = specificity, PPV = positive predictive value, NPV = negative predictive value.

Figure 1

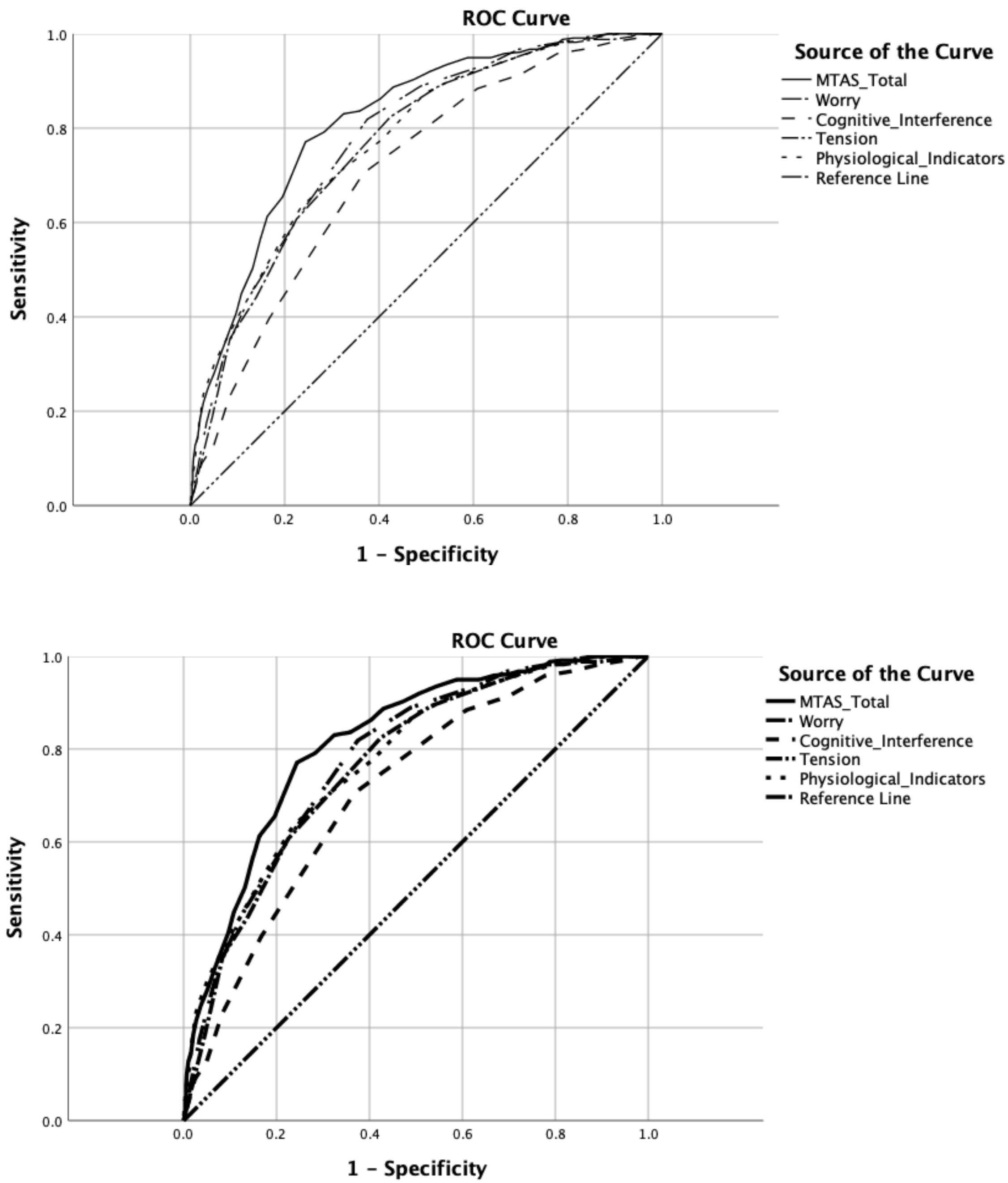
Hypothesized Structure of the Measurement Model



Note. λ_{G1} to λ_{G6} = factor loadings for generalized anxiety, λ_{P1} to λ_{P6} = factor loadings for panic, λ_{W1} to λ_{W9} = factor loadings for worry, λ_{CI1} to λ_{CI4} = factor loadings for cognitive interference, λ_{T1} to λ_{T4} = factor loadings for tension, λ_{PI1} to λ_{PI4} factor loadings for physiological indicators, and ε_1 to ε_{31} = residual variances. FSM = Free School Meals.

Figure 2

Receive Operator Characteristic Curve with Generalized Anxiety and Panic



Note. MTAS_Total = Multi dimensional Test Anxiety Scale total score.

Appendix: MTAS Items

1. Before a test/ exam, I am worried I will fail. (W)
2. I forget previously known material before taking a test/exam. (CI)
3. Even when I have prepared for a test/ exam I feel nervous about it. (T)
4. Before I take a test/ exam my hand trembles. (PI)
5. During tests/ exams, I worry about the consequences of failing. (W)
6. I forget facts I have learnt during tests/exams. (CI)
7. I feel tense before taking a test/exam. (T)
8. My heart races when I take a test/exam. (PI)
9. After a test/exam, I am worried I have failed. (W)
10. During tests/exams, I forget things that I have learnt (CI)
11. Just before I take a test/exam, I feel panicky. (T)
12. During a test/ exam I experience stomach discomfort.(PI)
13. During a test/ exam, I worry that I gave the wrong answers. (W)
14. During tests/exams, I find it hard to concentrate. (CI)
15. Before a test/exam, I feel nervous (T)
16. During a test/ exam, my muscles are tight. (PI)

W = Worry, CI = Cognitive Interference, T = Tension, and PI = Physiological Indicators.