2013

Assessing the construct validity of test anxiety: The influence of test characteristics and impact on test score criterion validity

Andrea Gaye-Valentine, *University at Albany, State University of New York*
Marcus Crede, *George Washington University*

TPM®

ASSESSING THE CONSTRUCT VALIDITY OF TEST ANXIETY: THE INFLUENCE OF TEST CHARACTERISTICS AND IMPACT ON TEST SCORE CRITERION VALIDITY

ANDREA GAYE-VALENTINE
UNIVERSITY AT ALBANY – STATE UNIVERSITY OF NEW YORK

MARCUS CREDÉ
GEORGE WASHINGTON UNIVERSITY

The construct validity of test anxiety requires not only confirmation of its factor structure, distinction from related affective constructs, and negative relationship with test performance, but also requires that the test scores of test anxious individuals have lower levels of predictive validity than the test scores of individuals with lower levels of test anxiety. This paper examines these varying components of the construct validity of test anxiety and presents evidence of the constructs' unique position within the nomological network as well as evidence that test anxiety reduces the predictive validity of scores on a measure of academic aptitude. Implications for assessment in high stakes settings are discussed.

Key words: Test anxiety; Construct validity; Moderation; Worry; Emotionality.

Correspondence concerning this article should be addressed to Marcus Credé, Department of Organizational Sciences and Communication, George Washington University, 600 21st Street NW, Washington, DC 20052, USA. Email: mac483883@gwu.edu

INTRODUCTION

Test anxiety is a widely studied construct in the educational literature because of its implication for the manner in which students' class-specific knowledge and skills are assessed. Students who score highly on self-report measures of test anxiety have repeatedly been shown to perform more poorly on classroom tests than students with lower levels of self-reported test anxiety (e.g., Hembree, 1988) — an effect widely assumed to be the result of the debilitating effect of the anxiety experienced during high stakes testing sessions on both the retrieval and processing of test relevant information (Meijer & Oostdam, 2007; Putwain, Connors, & Symes, 2010; Zeidner, 1998). This debilitating effect is not only relevant in educational settings but also in other settings in which test scores are used to make high stakes decisions, such as the college admissions process and employee selection in organizations. If test anxiety results in a "mismeasure" of test anxious individuals then it is also likely to increase the rate of incorrect admissions and hiring decisions since applicants who are test anxious but otherwise well-qualified are unlikely to be admitted or hired.

Given the possible negative impact of test anxiety on the quality of decisions made in high stakes settings we aim, in this paper, to examine the construct validity of test anxiety. In ad-

dition to a traditional examination of the structure of test anxiety, its convergent and discriminant validity, and the relationship between test anxiety and test performance (three issues that have been examined by previous research) we also focus on two additional issues that are central to the construct validity of test anxiety but that have not been widely examined in the test anxiety literature: the impact of test anxiety on the predictive validity of test scores, and the moderating impact of test characteristics on the test anxiety-test performance relationship. Both of these are important, and previously largely unexamined components of the construct validity of test anxiety but both also have potentially profound implications for testing practices in academic and organizational settings. We briefly discuss each of these components of the construct validity of test anxiety below.

The Structure of Test Anxiety

Perspectives on the nature and structure of test anxiety vary widely, but one relatively widely accepted view of the structure of test anxiety is that it is comprised of a cognitive component and an affective component (Spielberger & Vagg, 1995; Zeidner, 1998); a theoretical structure that is reflected in many widely used measures of test anxiety (e.g., Test Anxiety Questionnaire; Liebert & Morris, 1967), where the cognitive and affective components are often referred to as worry and emotionality, respectively (e.g., Morris, Davis, & Hutchings, 1981; Spielberger, 1980). Worry refers largely to cognitive responses to testing situations that include concern about performance prior to and after taking the test, and concerns about the consequences of poor performance. These concerns are themselves likely to be a function of preparation for the test, past experiences in evaluative situations as well as a preoccupation with failure and general low self-efficacy (Zeidner, 1998). Emotionality, on the other hand, primarily refers to emotional and physiological responses that occur while taking the test. Both worry and emotionality are thought to interfere with the recall and processing of test relevant information: worry because of the cognitive interference of test-irrelevant thoughts, and emotionality because of the associated difficulty of maintaining focus and accessing relevant information in memory (Spielberger & Vagg, 1995).

The distinction between worry and emotionality, confirmed in a variety of factor analytic examinations (e.g., Hocevar & El-Zahhar, 1988; Gierl & Rogers, 1996; Papantoniou, Moraitou, & Filippidou, 2011), is based on the theoretical distinction between affective and cognitive reactions to stressful situations but also requires that factor analytic findings are supportive of the worry-emotionality distinction and that worry and emotionality exhibit different correlational profiles with related constructs.

Hypothesis 1: worry and emotionality constitute two different factors of test anxiety.

Hypothesis 2: worry and emotionality exhibit unique correlational profiles with other variables in the nomological network.

Convergent and Discriminant Validity of Test Anxiety

Although the two-factor view of test anxiety is relatively widely accepted, the position of test anxiety within the nomological network is less clear. That is, the relationship of test anxiety

with other constructs has been widely examined, but few researchers have formally examined the degree to which scores on measures of test anxiety (reaction to specific test situation) can be distinguished from scores on measures of other constructs, particularly those that also reflect a more general tendency to experience negative affect (e.g., general anxiety), and those that reflect a general orientation toward tests and evaluation of tests (e.g., test attitudes).

Hypothesis 3: both the worry and emotionality components of test anxiety represent constructs that are significantly different from other negative affective constructs and from evaluations of the value of tests.

Main Effect on Test Performance

The relationship of test anxiety with test performance is likely the most widely studied relationship in the test anxiety literature, and meta-analytic evidence (Hembree, 1988) suggests negative relationships with both worry and emotionality. The worry component of test anxiety is thought to reflect anticipated performance or perceived competence on the test (Anderson & Sauser, 1995; Lang & Lang, 2010; Putwain & Symes, 2010; Putwain, Woods, & Symes, 2010), and since most test takers have some reasonably accurate sense of their preparation (pre-test) or actual performance (post-test), the worry component of test anxiety should be negatively correlated with test performance. This is the deficit model of test anxiety (Tobias, 1985) whereby test anxiety is the result (not cause) of poor academic performance. The theoretical relationship of the emotionality component of test anxiety with test performance is somewhat different. Emotionality refers to the experience of negative emotions such as panic, fear, and anxiety while taking the test; emotions that are likely to interfere with the recollection and processing of test relevant information. As such, emotionality is related to test performance because it reflects anxiety that directly interferes with test performance (e.g., Hembree, 1988; Meijer & Oostdam, 2007; Musch & Broder, 1999; Putwain, Connors, & Symes, 2010).

Hypothesis 4: test anxiety (worry and emotionality) is negatively related to test performance.

Moderating Role of Test Type

Although test anxiety appears to have a negative main effect on test performance, the strength of this effect may be moderated by test characteristics. Two test characteristics that may influence the strength of the effect are the time limits for a test and the degree to which the test follows a typical test format. Tests taken under limited time conditions (i.e., speed tests) are likely to result in a greater amount of experienced anxiety than tests taken in the absence of such time limitations (power tests). Similarly, tests that have the look and feel of standardized tests (e.g., questionnaire booklet, multiple choice answers) are more likely to induce anxiety than tests without these characteristics. For example, cognitive ability can be assessed using traditional test formats, or can be assessed using tools that have characteristics more often associated with games or puzzles than with tests. For example, components of the Wechsler Adult Intelligence Scales (e.g., Block Design test), which do not involve written questions or multiple-choice answers,

might therefore result in lower levels of test anxiety. Similar assessment tools that share characteristics with games and puzzles are widely used in the research literature to provide indicators of cognitive ability (e.g., Tower of Hanoi; Welsh, Satterlee-Cartmell, & Stine, 1999).

Hypothesis 5: the relationship between test anxiety and test scores is stronger for speed tests than for power tests.

Hypothesis 6: the relationship between test anxiety and test scores is stronger for traditional paper-and-pencil tests than for puzzle-format tests.

Moderating Effect Test Anxiety on Test Score Validity

The characteristics of test anxiety discussed above (factor structure, distinction from other negative affective constructs, and negative relationships with test scores), are necessary but not sufficient conditions for the construct validity of test anxiety. One additional required characteristic is that test anxiety interferes with the recall and processing of test relevant information. That is, test anxious individuals are thought to perform more poorly on tests than they would do in the absence of test anxiety. Curiously, this effect has not been widely examined in the test anxiety literature, perhaps because it cannot be illustrated simply by demonstrating that the scores on a measure of test anxiety negatively correlate with test scores. The experience of anxiety during tests may be due to 1) the inherent nature of the testing situation, 2) a lack of preparation for the test, or 3) the test taker's anticipated low performance on the test. In all three instances, the individual would report high levels of test anxiety and perform poorly on the test but only in the first of these cases would the interference effect of test anxiety be present. That is, the illustration that scores on a self-report measure of test anxiety correlate negatively with scores on a mathematics exam is not sufficient to illustrate that test anxiety has interfered with the recall and processing of test relevant information.

What would constitute evidence of the interfering effect of test anxiety is if test scores for test anxious individuals are less valid than scores on the same test for individuals with lower levels of test anxiety — an effect proposed (but tested only with simulations) by Reeve, Heggested, and Lievens (2009). For example, in a high stakes setting such as selection for a job the relationship between scores on a cognitive ability test and future job performance should be weaker for applicants with high levels of test anxiety than for those with low levels of test anxiety. Those applicants reporting high levels of test anxiety will, on average, score lower on the cognitive ability test than those reporting low levels of test anxiety; but some of these individuals will have significantly higher levels of cognitive ability than indicated by their test score. The result is that test scores for these individuals will not accurately predict their future job performance (not itself influenced by test anxiety), thus lowering the predictive validity of cognitive ability test scores for the overall group of test anxious applicants. Such an effect would suggest that selection systems in organizations and educational institutions may need to reduce their reliance on tests that are likely to be affected by test anxiety. Given that the emotionality component of test anxiety refers to the experience of interfering negative emotions while taking the test, we anticipate that this effect is likely to be stronger for the emotionality component of test anxiety than for the worry component of test anxiety which primarily references worry about anticipated test performance prior to the test and after having completed the test.

Hypothesis 7a: test anxiety moderates the predictive validity of scores on tests taken under high stakes conditions.

Hypothesis 7b: the moderating effect of test anxiety on test score validity is stronger for the emotionality component of test anxiety than for the worry component of test anxiety.

The Present Study

In addition to examining two components of the construct validity of test anxiety that have not been widely examined in the literature (i.e., the effect of test type and the effect of test anxiety on test score validity), our paper also makes a methodological contribution to the test anxiety literature by reducing the influence of a number of important confounding variables in prior examinations of test anxiety.

Examinations of the influence of test anxiety on test scores often take place in classroom settings in which students provide self-reports on their level of test anxiety while taking an in-class examination. Indeed, Hembree's (1988) meta-analytic review of the test anxiety literature summarized data from 176 such studies. Such research designs are however vulnerable to assessing anxiety that is not necessarily a function of being tested but rather due to a lack of preparation for the test.

In this paper, we examine performance on four tests that are not subject to preparation in the same manner that most in-class examinations and tests are: the SAT test; a standardized university admission test that is widely used in the United States of America, and three measures of cognitive ability. Test preparation courses are widely thought to improve performance on tests such as those used in admissions or selection contexts but such effects appear to generally be weak (e.g., Kulik, Bangert-Drowns, & Kulik, 1984; Ryan, Ployhart, Greguras, & Schmit, 1998), and are certainly much weaker than the effect of studying and/or revising on academic performance (Credé & Kuncel, 2008). Further, our focus on cognitive ability tests also allows us to examine the moderating influence of test anxiety on test score validity by allowing a contrast of performance on a high stakes test (the SAT) with performance on a fundamentally similar test (i.e., the Wonderlic Personnel Test) taken under low stakes (i.e., lab) conditions. Such an examination requires a criterion variable that is not also subject to influence by test anxiety. Thus, college grade point average (GPA; the average of grades received in university classes) would not be a suitable criterion since it is itself a function of test scores taken under high stakes conditions.

METHOD

Participants

Undergraduate students were recruited from the psychology participant pool of a large U.S. public university. Participants signed up as potential participants based on a short description of the purpose of the study although this short description did not include mention of the fact that participants would be asked to complete the tests described below in order to minimize the degree to which test anxious individuals avoided participation in the study. At the beginning of

the actual data collection, potential participants were fully informed about the nature of the tests they would be asked to complete (as part of the informed consent procedure) and were given an opportunity to decline participation at this point without penalty (none declined participation). All 199 participants received course credit for their participation as well as a small ($10) monetary incentive for test performance. Participants were predominantly female (64.3%) and Caucasian (64.8%) with an average age of 19.16 years (*SD* = 2.35).

Measures

Wonderlic Personnel Test. The 50-item Wonderlic Personnel Test (WPT; Wonderlic, 1970) is a 12-minute test of learning and problem solving ability used widely in organizational settings and was used as a speed test measure of cognitive ability. The test-retest reliability of Wonderlic scores has been found to be very high (e.g., $r = .94$; Dodrill, 1983).

Raven's Progressive Matrices. The Advanced Raven's Progressive Matrices (RPM; Raven, Raven, & Court, 2003) is a test of abstract reasoning used in organizational and educational settings as a measure of cognitive ability; it has been found to produce scores that are highly correlated with scores on measures of cognitive ability (e.g., $r = .74$ with scores on Wechsler Adult Intelligence Scales; McLaurin, Jenkins, Farrar, & Rumore, 1973) and exhibits high test-retest reliability ($r = .83$; Bors & Forrin, 1995). Items take the form of a pattern completion task; each item is comprised of a 3×3 matrix of shapes one of which is missing, and respondents are required to select the shape (from six possible shapes) that would match the pattern of the other eight shapes. Participants were given unlimited time to complete 12 items of increasing difficulty. The RPM was used as a power test measure of cognitive ability.

Tower of Hanoi. The Tower of Hanoi (TOH) is a puzzle task, completion of which has been found to require cognitive skills (e.g., Zook, Davalos, DeLosh, & Davis, 2004). The puzzle is comprised of a series of disks varying in sizes and three rods with the goal being to transfer the entire set of disks from a peg on the left to the peg on the right. The disks can be moved to any other peg but a larger disk cannot be placed on top of a smaller one, and only one disk can be moved at a time. Participants start off with three disks and are given more disks to transfer as they progress through the task. The highest number of sets of disks the participant was able to complete was used as the TOH score. Mildren, Cazzell, and Holland (2012) report internal consistency estimates for different tasks using the TOH as alpha = .69.

SAT Scores. Self-report SAT scores were used as an indicator of cognitive ability obtained in a high stakes testing situation. Prior meta-analytic research has shown that self-reported SAT scores are highly correlated with actual SAT scores ($r = .82$; Kuncel, Credé, & Thomas, 2005) and that SAT scores are highly correlated with scores on measures of cognitive ability (Coyle & Pillow, 2008). SAT scores have also exhibited excellent reliability in prior studies (e.g., KR-20 reliability > .90; Mattern, Camara, & Kobrin, 2007).

Test Anxiety Inventory. The 20-item Test Anxiety Inventory (TAI) was used to assess an individual's level of test anxiety (Spielberger, 1980). The inventory is comprised of an eight-item worry subscale and an eight-item emotionality subscale. Each statement is answered on a 4-point Likert scale (1 = *almost never* to 4 = *almost always*).

Test Attitude Survey. The Test Attitude Survey (TAS) was used to measure the attitudes of individuals towards tests (Sarason, 1978). The TAS consists of nine subscales; however an exploratory factor analysis suggested only two factors: positive attitudes and negative attitudes. Each statement is answered on a 5-point Likert scale (1 = *very inaccurate* to 5 = *very accurate*).

Negative Affective Constructs. General trait anxiety (24 items), neuroticism (20 items), and depression (10 items) were assessed using scales from the International Personality Item Pool (Goldberg, 1999). Each statement is answered on a 5-point Likert scale (1 = *very inaccurate* to 5 = *very accurate*).

RESULTS

The correlations among all examined variables are presented in Table 1. The reliability estimates (presented along the diagonal) ranged from .66 to .93.

Structure of Test Anxiety

The worry and emotionality components of test anxiety exhibited a correlation of $r = .78$, suggesting nontrivial overlap. Confirmatory factor analysis (CFA) was therefore used to compare the fit of a single-factor solution with the fit of the hypothesized two-factor solution. CFA assumes that data is interval in nature and that latent factors can account for the covariation among manifest variables (i.e., among scores on items), both assumptions that are reasonable in this situation. To interpret results we relied on the recommendations by Hu and Bentler (1999) that comparative fit index (CFI) and incremental fit index (IFI) values be greater than .95 to indicate good fit and that a RMSEA value close to or less than .06 indicates good fit. Chi-square values and Akaike information criterion (AIC) values are used primarily for model comparison. Results showed that a two-factor solution (worry and emotionality) exhibited good fit in an absolute sense (RMSEA = .065, CFI = .98, IFI = .98, AIC = 254.94) and a significantly ($p < .01$) better fit ($\Delta\chi^2 = 97.27$, $\Delta df = 1$) than a single factor solution (RMSEA = .094, CFI = .97, IFI = .97, AIC = 350.21). The latent factors in the two-factor solution were correlated at $r = .88$. The results of the CFA provide support for Hypothesis 1 that worry and emotionality constitute two meaningfully different factors of test anxiety.

Discriminant Validity of Worry and Emotionality

CFA was also used to examine the discriminant validity of worry and emotionality from related negative affective constructs. To do this we first examined which of the other measured variables exhibited the strongest relationships with worry and emotionality. The largest correlate of the worry subscale was neuroticism while the largest correlate of the emotionality subscale was general trait anxiety. We then compared the fit of two single-factor models (worry and neuroticism items loading onto one latent variable and emotionality and trait anxiety items loading

TABLE 1
Correlation matrix of study variables

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Test Anxiety: Worry	2.02	0.69	.89											
2. Test Anxiety: Emotionality	2.31	0.73	.78	.90										
3. Test Anxiety: Total	2.19	0.67	.92	.95	.95									
4. Depression	2.20	0.62	.46	.35	.42	.86								
5. General Anxiety	2.94	0.60	.50	.57	.59	.67	.93							
6. Neuroticism	2.61	0.58	.54	.56	.59	.80	.89	.90						
7. Test Attitudes: Negative	2.60	0.35	.49	.40	.48	.47	.42	.50	.66					
8. Test Attitudes: Positive	3.24	0.30	.02	.18	.13	–.06	.11	.01	.09	.68				
9. Raven's Progressive Matrices (RPM)	5.77	2.45	–.13	–.16	–.15	.12	.03	.05	–.18	–.05				
10. Wonderlic Personnel Test (WPT)	21.63	5.07	–.30	–.19	–.26	–.07	–.20	–.16	–.41	–.04	.35			
11. Tower of Hanoi (TOH)	5.16	0.72	.00	–.04	–.04	–.03	–.14	–.08	–.10	.03	.15	.22		
12. SAT	1107.96	181.79	–.23	–.25	–.26	–.05	–.24	–.21	–.18	–.11	.25	.51	.09	

Note. Numbers on the diagonal are local reliability estimates.

onto one latent factor) with two models in which the items of the respective constructs loaded onto separate constructs. The fit of a single-factor model for the worry and neuroticism items was poor (RMSEA = .15, CFI = .87, IFI = .87, NNFI = .87, AIC = 1944.55) while a two-factor solution was statistically better ($\Delta\chi^2 = 662.56$, $\Delta df = 1$, $p < .01$), and also better in an absolute sense (RMSEA = .11, CFI = .91, IFI = .91, AIC = 1169.99). The latent factor intercorrelation was r = .59. Similarly, the fit of a single-factor model for the emotionality and general anxiety items was not good (RMSEA = .12, CFI = .91, IFI = .91, AIC = 1997.54), while the fit of a two-factor solution was good (RMSEA = .073, CFI = .95, IFI = .95, AIC = 1083.16) in absolute sense and statistically significantly better ($\Delta\chi^2 = 914.38$, $\Delta df = 1$, $p < .01$). The latent factor intercorrelation was r = .64. The results of the CFA analyses provide support for Hypothesis 3 that both worry and emotionality are significantly different from other highly correlated variables.

Relationship with Other Variables

Using the Meng, Rosenthal, and Rubin (1992) method, correlated correlation coefficients were compared to examine if worry and emotionality exhibit different relationships with related variables. Results (Table 2) illustrate significantly different correlations of worry and emotionality for four of the examined correlates. Three correlates (i.e., depression, negative test attitudes, and WPT) were more highly correlated with the worry component of test anxiety than with the emotionality component of test anxiety while two correlates (i.e., general anxiety and positive test attitudes) were more highly correlated with the emotionality component. These results lend support to Hypothesis 2 in that worry and emotionality exhibit different correlational profiles with related variables.

TABLE 2
Comparing correlation coefficients

Y	r_{worry}	$r_{emotionality}$	Z
Depression	.46	.35	2.47*
General anxiety	.50	.57	–1.69*
Neuroticism	.54	.56	–0.31
Test Attitudes: Negative	.49	.40	2.17*
Test Attitudes: Positive	.02	.18	–3.35*
RPM	–.13	–.16	0.68
WPT	–.30	–.19	–2.40*
TOH	.00	–.04	0.88
SAT	–.23	–.25	0.30

Note. * = $p < .05$; r_{worry} = correlation with worry component of test anxiety; $r_{emotionality}$ = correlation with emotionality component of test anxiety; Z = Meng and Rosenthal test of difference between correlated correlations. RPM = Raven's Progressive Matrices; WPT = Wonderlic Personnel Test; TOH = Tower of Hanoi.

Test Anxiety and Performance Outcomes

Table 1 presents the Pearson correlations between test anxiety and the measures of cognitive ability and academic performance. Hypothesis 4 is supported with significant ($\alpha = .05$) negative correlations between test anxiety and the RPM scores ($r = -.15$), WPT scores ($r = -.26$), and SAT scores ($r = -.26$). Although these correlations are only weak to moderate (Cohen, 1988), they are in the expected direction and in line with previous studies examining cognitive ability scores and test anxiety (Hembree, 1988). TOH scores ($r = -.04$) were not significantly related to test anxiety. The relative strength of the relationships of test anxiety with the three directly assessed measures of cognitive ability was in line with the order of strength implied by Hypotheses 5 and 6. In order to examine Hypothesis 5 and Hypothesis 6 more formally, we used the Meng et al. (1992) procedure to compare the relationships of test anxiety with the three indicators of cognitive ability directly assessed in this study (WPT, RPM, and TOH). The test anxiety-WPT relationship was found not to be significantly stronger than the test anxiety-RPM relationship ($Z = 1.40$, $p = .08$), although the difference was relatively large ($r = -.26$ vs. $r = -.15$). The test anxiety-WPT relationship was found to be significantly stronger than the test anxiety-TOH relationship ($Z = 2.25$, $p < .05$), while the test anxiety-RPM relationship was not significantly stronger than the test anxiety-TOH relationship ($Z = 1.00$), lending partial support for Hypothesis 6 but not for Hypothesis 5.

Moderating Effect of Test Anxiety on Test Score Validity

Moderated regression analysis was used to examine whether the relationship between test scores from a high stakes assessment setting (SAT scores) and scores from a low stakes assessment setting (cognitive ability scores) are moderated by overall test anxiety and either the worry or emotionality components of test anxiety (Hypotheses 7a and 7b). Results for the test most similar to the SAT in terms of format and time limit (i.e., the WPT) indicate that overall test anxiety significantly moderates the SAT-WPT relationship ($\Delta R^2 = .032$, $\Delta F = 4.46$, $p = .04$). Further, this effect was even stronger for emotionality ($\Delta R^2 = .048$, $\Delta F = 6.79$, $p = .01$), and not significant for worry ($\Delta R^2 = .014$, $\Delta F = 1.89$, $p = .17$). Figure 1 illustrates that the effect for emotionality is strong and in the expected direction. That is, the relationship between SAT scores and WPT scores is significantly stronger for individuals with low levels of emotionality than for individuals with high levels of emotionality. For participants in the top 50% of emotionality scores the SAT-WPT relationship was $r = .37$, while for participants in the bottom 50% of emotionality scores the SAT-WPT relationship was $r = .65$.

Discussion

Test anxiety occupies a relatively unique position in the nomological network. As is the case for other constructs, issues of factor structure and relationships with theoretically related variables (i.e., convergent, discriminant, and predictive validity) are important components of the

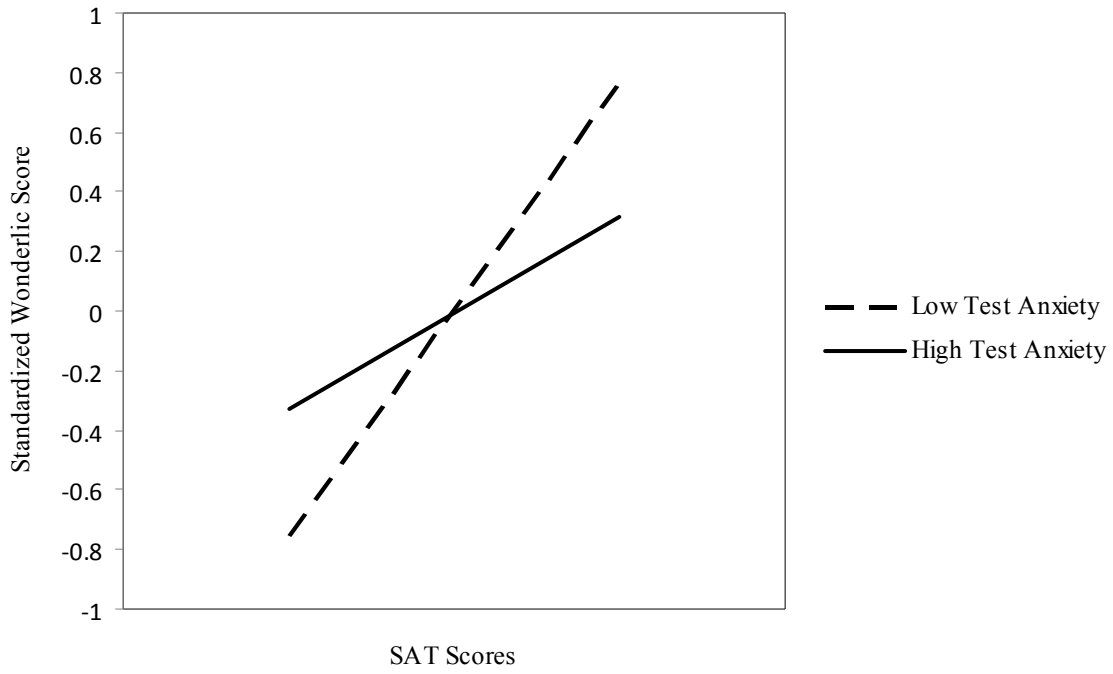

FIGURE 1
Moderating effect of test anxiety (emotionality) on the relationship between scores on SAT and WPT.

construct validity of test anxiety. One additional component — and in many ways the defining component — is the theoretical effect that test anxiety has on the meaning attached to scores on tests of knowledge, skills, aptitude, and ability, tests that include those used in high stakes settings such as college admissions and employee selection settings. The aim of this paper was to examine the construct validity of test anxiety in a manner that also allowed examination of this defining component of the test anxiety construct, a component that, to our knowledge, has never previously been examined.

Our results are broadly supportive of the construct validity of test anxiety. Our factor analytic and correlational results support a two-dimensional view of test anxiety that distinguishes between the worry and emotionality components of the construct. The relationships of test anxiety with scores on tests of cognitive ability are also largely as expected, with high test anxiety scores being related to lower performance on cognitive ability tests, particularly for those that are taken under time constraints.

Most important are our results that test anxiety substantially moderates the relationship between scores on a test taken under high stakes conditions and scores on a similar test taken under low stake conditions. This finding strongly suggests that at least some test anxious individuals are being significantly "mismeasured" in high stakes testing situations, and that their scores are therefore less predictive of future performance. This, in turn, has important implications for testing in high stakes settings. Colleges and employers rely on scores on tests such as the SAT and WPT to improve the quality of admissions and hiring decisions respectively. The underlying assumption is that individuals who score highly on tests such as the SAT or WPT are more likely

to succeed in college or at work, an assumption supported by decades of research (Kuncel, Hezlett, & Ones, 2004; Schmidt & Hunter, 1998). Our results suggest though that the test scores of some individuals reporting high levels of test anxiety are less predictive of future performance than for those reporting low levels of test anxiety; therefore, the overall validity of selection systems that rely on tests that are likely to cause test anxiety (e.g., speed tests) is lower than it would be if the selection system relied less on anxiety inducing tests and more on alternatives such as power tests or noncognitive predictors.

Although the implications for high-stakes test settings are relatively clear (some test anxious individuals are being mismeasured), the remedy is less clear. One possible way of reducing the effect of test anxiety on the accuracy of selection decisions might be to rely less on speed tests and more on power tests that are not characterized by strict time limits. However, relying on untimed tests might present practical difficulties for the types of massed testing sessions that characterize many college admissions or employee selection settings. Another possibility is to reduce the degree to which admissions or hiring decisions are based on scores from tests taken under high stakes settings. In employment settings this may involve a greater reliance on interviews, work samples, or personality and integrity tests that are less prone to test anxiety but that are also predictive of job performance. Similarly, college admissions decisions may reduce the emphasis on SAT scores and rely more on alternate valid predictors of academic achievement that are perhaps less influenced by test anxiety such as high school grades. The most significant challenge is identifying alternate predictors of college achievement that are not prone to faking or socially desirable responding while simultaneously not sacrificing the very substantial predictive validity provided by SAT scores for the majority of students.

Limitations and Future Research

If, as our results suggest, test anxiety reduces the validity of cognitive ability test scores then this decrement in validity should be reflected in a decreased relationship (for those reporting high levels of test anxiety) between cognitive ability test scores and important outcomes such as academic performance and job performance. Unfortunately, academic performance in college is itself primarily measured using high stakes assessment practices. As such this specific moderating effect of test anxiety on the predictive validity of cognitive ability test scores is difficult to examine in academic settings since both the predictor (e.g., SAT scores) and the criterion (e.g., class grades) are equally affected by test anxiety. We therefore propose that future research in this area focuses on work settings since these are typically settings where performance is not assessed by multiple-choice tests taken under strict time limitations, but rather by supervisors' evaluations of everyday behavior.

We also hope that future research is better able to distinguish between individuals who experience anxiety because they are inadequately prepared or because they have prior knowledge of how they are likely to do on the test and those who experience anxiety primarily due to the testing situation. This may require treating test preparation and prior test experience as a covariate in research designs or relying on subjects who have not previously been tested on the examined ability or skill.

Our measurement of test anxiety, and to a lesser degree cognitive ability and SAT scores, were characterized by measurement error which would have attenuated the observed moderation

effect downward. As such the impact of test anxiety on test validity is likely even stronger than observed in this paper.

CONCLUSION

Our results confirm the two-factor structure of test anxiety, its relationship to test performance, and its distinction from related affective constructs. Most importantly, this study is perhaps the first study to illustrate that test anxiety may significantly reduce the quality of the inferences that researchers and practitioners can draw from the test results of test anxious individuals.

REFERENCES

Anderson, S. B., & Sauser, W. I. (1995). Measurement of test anxiety: An overview. In C. D. Spielberger (Ed.), *Test anxiety: Theory, assessment and treatment* (pp. 15-33). Washington, DC: Taylor-Francis & Co.

Bors, D. A., & Forrin, B. (1995). Age, speed of information processing, recall, and fluid intelligence. *Intelligence*, *20*, 229-248. doi:org/10.1016/0160-2896(95)90009-8

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Coyle, T., & Pillow, D. (2008). SAT and ACT predict college GPA after removing "g". *Intelligence*, *36*, 719-729. doi:org/10.1016/j.intell.2008.05.001

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, *3*, 425-453. doi:org/10.1111/j.1745-6924.2008.00089.x

Dodrill, C. (1983). Long-term reliability of the Wonderlic Personnel Test. *Journal of Consulting and Clinical Psychology*, *51*, 316-317. doi:org/10.1037/0022-006X.51.2.316

Gierl, M. J., & Rogers, W. T. (1996). A confirmatory factor analysis of the Test Anxiety Inventory using Canadian high school students. *Educational and Psychological Measurement*, *56*, 315-324. doi:org/10.1177/0013164496056002012

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

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

Hocevar, D., & El-Zahhar, N. E. (1988). Arousability, trait anxiety, and the worry and emotionality components of test anxiety. *Anxiety Research*, *1*, 99-113. doi:org/10.1080/10615808808248223

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

Kulik, J. A., Bangert-Drowns R. L., & Kulik, C. C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin*, *95*, 179-188. doi:org/10.1037/0033-2909.95.2.179

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, *75*, 63-82. doi:org/10.3102/00346543075001063

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148-161. doi:org/10.1037/0022-3514.86.1.148

Lang, J. W. B., & Lang, J. (2010). Priming competence diminishes the link between cognitive test anxiety and test performance: Implications for the interpretation of test scores. *Psychological Science*, *21*, 811-816. doi:org/10.1177/0956797610369492

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

Mattern, K., Camara, W., & Kobrin, J. L. (2007). *SAT writing: An overview of research and psychometrics to date* (College Board Research Report RN-20). New York, NY: The College Board.

McLaurin, W., Jenkins, J., Farrar, W., & Rumore, M. (1973). Correlations of IQ on verbal and non-verbal test of intelligence. *Psychological Reports*, *33*, 821-822. doi:org/10.2466/pr0.1973.33.3.821

Meijer, J., & Oostdam, R. (2007). Test anxiety and intelligence testing: A closer examination of the stage-fright hypothesis and the influence of stressful instruction. *Anxiety, Stress & Coping: An International Journal*, *20*, 77-91.

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172-175. doi:org/10.1037/0033-2909.111.1.172

Mildren, W. G., Cazzell, J. D., & Holland, M. A. (2012). *Surprise findings from Tower of Hanoi research on executive cognitive function in older adults: Assess, intervene, or both?* Paper presented at the 23rd International Nursing Research Congress. Brisbane, Australia.

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology*, *73*, 541-555. doi:org/10.1037/0022-0663.73.4.541

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

Papantoniou, G., Moraitou, D., & Filippidou, D. (2011). Psychometric versions of the Greek version of the Test Anxiety Inventory. *Psychology*, *2*, 241-247. doi:org/10.4236/psych.2011.23038

Putwain, D. W., Connors, L., & Symes, W. (2010). Do cognitive distortions mediate the test anxiety-examination performance relationship? *Educational Psychology*, *30*, 11-26. doi:org/10.1080/01443410903328866

Putwain, D. W., & Symes, W. (2010). Achievement goals as mediators of the relationship between competence beliefs and test anxiety. *British Journal of Educational Psychology*, *82*, 207-224. doi:org/10.1111/j.2044-8279.2011.02021.x

Putwain, D. W., Woods, K. A., & Symes, W. (2010). Personal and situational predictors of test anxiety of students in post-secondary education. *British Journal of Educational Psychology*, *80*, 137-160. doi:org/10.1348/000709909X466082

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and vocabulary scales.* San Antonio, TX: Harcourt Assessment.

Ryan, A. M., Ployhart, R. E., Greguras, G. J., & Schmit, M. J. (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. *Personnel Psychology*, *51*, 599-621. doi:org/10.1111/j.1744-6570.1998.tb00253.x

Reeve, C. L., Heggested, E. D., & Lievens, F. (2009). Modeling the impact of test anxiety and test familiarity on the criterion-related validity of cognitive ability tests. *Intelligence*, *37*, 34-41. doi:org/10.1016/j.intell.2008.05.003

Sarason, I. G. (1978). The Test Anxiety Scale: Concept and research. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and anxiety* (Vol. 5, pp. 193-216). Washington, DC.: Hemisphere Publishing Corp.

Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274. doi: org/10.1037/0033-2909.124.2.262

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

Spielberger, C. D., & Vagg, P. R. (1995). Test anxiety: A transactional model. In C. D. Spielberger (Ed.), *Test anxiety: Theory, assessment and treatment* (pp. 3-14). Washington, DC: Taylor-Francis & Co.

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

Welsh, M. C., Satterlee-Cartmell, T., & Stine, S. (1999). Towers of Hanoi and London: Contribution to working memory and inhibition to performance. *Brain and Cognition*, *40*, 231-242. doi:org/10.1006/brcg.1999.1123

Wonderlic, E. F. (1970). *Wonderlic Personnel Test Manual*. Vernon Hills, Il: E.F. Wonderlic & Associates.

Zeidner, M. (1998). *Test Anxiety: The state of the art.* New York, NY: Plenum.

Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition*, *56*, 286-292. doi:org/10.1016/j.bandc.2004.07.003