Selection Test Anxiety: Exploring Tension and Fear of Failure Across the Sexes in Simulated Selection Scenarios

Julie M. McCarthy*
University of Toronto

Richard D. Goffin
University of Western Ontario

Standardized tests are commonly used to select individuals in both pre-employment and educational settings. Nevertheless, research has yet to sufficiently explore the relationship between test anxiety (TA) and test performance in selection contexts. The goal of this study was to assess the dimensionality of TA, the relations between TA and test performance, and the impact of gender on TA within the realm of personnel selection. Test anxiety was found to be bi-dimensional and negatively related to performance on a cognitively based selection instrument. Gender moderated the relation between TA and test performance, with stronger associations for males. This finding is consistent with the *sex-linked anxiety coping theory* that is advanced here. Implications of this study and avenues for future research are discussed.

Introduction

A wealth of research indicates that standardized tests can be an invaluable aid in identifying the most promising candidates for a job (Schmidt & Hunter, 1998), as well as the most promising students for an academic program (Kuncel, Hezlett, & Ones, 2001). Unfortunately, such tests commonly induce a phenomenon known as *test anxiety* (TA; Zeidner, 1998). This is problematic, as high levels of TA may have detrimental effects on selection test performance, resulting in scores that do not accurately reflect an individual's true levels of the respective attributes (Arvey, Strickland, Drauden, & Martin, 1990). This may ultimately result in the selection of less promising candidates (Arvey *et al*., 1990; Spielberger & Vagg, 1995). Moreover, gender differences in TA may influence selection test scores (Hembree, 1988) and ultimately result in the occurrence of adverse impact.

The current investigation examined the role of TA in selection contexts. There were three main goals, the first of which was to investigate the dimensionality of TA. Second, the relations between TA and "applicant" performance on a cognitively based selection test were explored. Third, the role of gender in selection TA was addressed and a *sex-linked anxiety coping theory* was advanced and tested. Consistent with this theory, test-taker gender was considered as a possible moderator of TA-test performance relations.

The construct of anxiety has been a prominent focus of attention in many theories of personality and abnormal psychology (Spielberger, 1972). Over the years, it has been used to refer to a stimulus, response, state, drive, and a trait (Endler, Edwards, & Vitelli, 1991). There appears to be a general consensus, however, that the more specific construct of TA is reflective of a situation-specific trait (Spielberger & Vagg, 1995). According to this perspective, writing a test can be perceived as an anxiety-provoking situation, and levels of trait anxiety specific to this situation can be assessed. Ultimately, TA can be defined as feelings of nervousness or apprehension that are stable within individuals across testing situations and are cognitive, behavioral, or affective in nature (Zeidner, 1998). Consistent with this definition, TA has been empirically linked to the personality trait of neuroticism (Sarason, 1959). This is not surprising, as neurotic individuals exhibit high levels of worry and emotionality – characteristics that form the basis of TA (Dobson, 2000). Moreover, general anxiety is typically regarded as one of the six primary components of neuroticism (Costa & McCrae, 1992).

*Address for correspondence: Julie M. McCarthy, Department of Management and Joseph L. Rotman School of Management, The University of Toronto, 1265 Military Trail, Toronto, Canada, M1C 1A4. E-mail: mccarthy@utsc.utoronto.ca

Dimensionality of Test Anxiety

As mentioned, the first goal of the current investigation was to assess the dimensionality of TA in a selection context. The literature pertaining to TA is enormous, with studies dating as far back as 48 years (i.e., Mandler & Sarason, 1952). Indeed, TA has been cited as one of the most salient constructs in the field of psychology (Zeidner, 1998). Much of the work in this area has been conducted in educational contexts, and has focused on determining the dimensionality of the TA construct (for reviews, see Hembree, 1998; Seipp, 1991; Spielberger & Vagg, 1995; Zeidner, 1998). In terms of prevalence, normative data suggest that in North America, between 30% and 52% of college students experience TA "often" or "almost always" (Spielberger, 1980). Additionally, Nottelmann and Hill (1977) found that TA was problematic for approximately 25–35% of school age children.

Early conceptualizations of TA were global in nature, encompassing reactions and attitudes towards school, courses, and instructors, as well as tests (Spielberger & Vagg, 1995). This is reflected in the first widely used TA scale, the Test Anxiety Questionnaire (Mandler & Sarason, 1952). This instrument contains 36 items and the developers report a reliability of .91 (Mandler & Sarason, 1952). It is noteworthy that although Mandler and Sarason implicitly conceptualized TA as unidimensional, they did not explicitly test the dimensionality of their measure. In 1967, Liebert and Morris proposed that TA consists of two major components: worry and emotionality. Worry was conceptualized as "concern" about the outcome of a test (i.e., fear of failure), while emotionality reflected the "autonomic arousal" experienced as a result of the testing situation (i.e., bodily tension). At least three instruments were developed to assess the worry and emotionality components – the Worry-Emotionality Questionnaire (Liebert & Morris, 1967), the Test Anxiety Inventory (TAI; Spielberger, 1980), and the Revised Worry-Emotionality Scale (Morris, Davis, & Hutchings, 1981). In recent years, a four-dimensional model of TA has been proposed by Sarason (1984) and operationalized via his 40-item Reactions to Tests questionnaire (RTT). The four components are tension (feeling nervous before and during a test), bodily symptoms (experiencing headaches and other unpleasant symptoms before and during a test), worry (thoughts of doing poorly on the test), and test-irrelevant thoughts (inability to stay focused on the test).

Presently, there is widespread consensus that TA is multidimensional in nature (Hembree, 1988). However, there is some lack of agreement over whether the two-dimensional or four-dimensional structure is more appropriate. Although Sarasons' four-dimensional structure has the potential to provide more comprehensive measurement (Elliot & McGregor, 1999), Liebert and Morris' two-dimensional structure appears to be superior for at least three reasons. First, the vast majority of studies on TA have found support for the two-factor structure (Cassady & Johnson, 2002), rendering the two-dimensional model the normative view (Spielberger & Vagg, 1995). Indeed, a meta-analysis of the TA literature provided separate estimates for the emotionality and worry components of TA because these two factors were included in "nearly all" of the questionnaires contained in the study (Seipp, 1991). Second, research on the four-dimensional structure indicates that the subscale intercorrelations are relatively large in magnitude, ranging as high as .90 (Kalechstein, Hocevar, Zimmer, & Kalechstein, 1989). This calls into question the proposition that there are more than two distinct factors of TA. Third, a careful examination of the four subscales proposed by Sarason (1984) reveals that conceptually, they can be collapsed into the worry and emotionality components. To be specific, the worry and test-irrelevant thoughts scales are reflective of the worry dimension, as they are cognitive in nature and focus on an individuals thoughts during a stressful encounter. In contrast, the remaining two scales, tension and bodily symptoms, are reflective of the emotionality dimension, as they represent the activation of the autonomic nervous system as a response to stressful situations (i.e., feeling jittery, experiencing headaches).

Unfortunately, past research on the dimensionality of TA has consisted of analyses wherein only one TA measure was examined at a time and exploratory factor analysis was largely relied upon (e.g., Liebert & Morris, 1967; Sarason, 1984). Although this research was valuable, it is difficult to draw firm conclusions as to the number and nature of factors underlying a construct without a broad representative sample of items from a number of TA measures. Conducting factor analyses on only one TA measure at a time may limit one's conclusions to that particular scale, as opposed to the overall construct. Thus, it would be more informative to conduct a factor analysis on the collective set of items taken from several TA measures that have been chosen to reflect the diversity of available instruments in this domain. Moreover, although past research has relied upon exploratory factor analysis, it is well known that confirmatory factor analysis is superior in terms of allowing firmer conclusions to be drawn regarding the dimensionality of an item pool. The current study advances past efforts by conducting *confirmatory* factor analyses on the *collective pool* of items from one-factor, two-factor, and four-factor TA instruments in a personnel selection context. It is important for readers to note that in cases where multiple measures can not be administered to the same sample, combining meta-analysis and confirmatory factor analysis is a valuable method that can be used to explicate the construct of interest (see Viswesvaran & Ones, 1995).

As discussed earlier, because the two-factor model has achieved the most support in the educational realm, there is some reason to expect that it will fit our data. However, to our knowledge this is the first investigation of the dimensionality of TA within a personnel selection scenario,

and it is quite possible that the unique stressors of a selection-testing situation will result in a TA dimensionality that differs from that of educational settings. Thus, we proposed to test one-factor, two-factor and four-factor models in order to evaluate the following research question:

Q1: Within personnel selection scenarios, is TA best represented by a one-factor, two-factor, or four-factor structure?

Studies of Test Anxiety and Test Performance

Once the appropriate dimensionality of TA was established as a result of testing Q1, the second goal of the current investigation was to examine the relation between TA and selection test performance. A considerable amount of research, much of which has been conducted in the educational realm, has found that high levels of TA are linked to low scores on tests (Zeidner, 1998). This is illustrated in a meta-analysis conducted by Hembree (1988), which was based on 562 studies. Measures of performance included cognitive ability tests, problem-solving tasks, achievement measures, and grades. Overall, TA demonstrated a weak negative relationship with test performance (uncorrected $\bar{r} = -.18$). A second meta-analysis, conducted by Seipp (1991), was based on 114 studies. In contrast to Hembree, Seipp (1991) restricted her focus to studies that investigated the relation between TA and academic test performance. Overall, Seipp's findings also revealed a weak negative relation (corrected $\bar{r} = -.23$).

The aforementioned investigations have relevance to selection contexts, where standardized testing is commonly used to select students for academic programs (Kuncel *et al*., 2001), and applicants for jobs (Morgan & Smith, 1996). In fact, analogous to tests used to assess students, selection instruments often assess components of cognitive ability (i.e., verbal ability, quantitative ability, reasoning; Hunter, 1986). We therefore find it surprising that only a handful of studies have examined the role of TA in selection testing contexts. To the best of our knowledge, the first investigation in this area was conducted as recently as 1990 and examined 263 applicants for a financial worker position (Arvey *et al*., 1990). Results indicated that TA was negatively related to scores on a cognitive ability test (uncorrected $\bar{r} = -.35$). Schmit and Ryan (1992) also examined TA in the realm of personnel selection by assessing 157 students using an employee selection simulation. Consistent with the findings of Arvey *et al*. (1990), a negative relation (uncorrected $\bar{r} = -.44$) was found between TA and the School and College Ability Test (SCAT). In a later study, Schmit and Ryan (1997) examined levels of TA in 3,290 police force applicants. Once again, a significant but weak negative relation (uncorrected $\bar{r} = -.11$) was obtained between TA and scores on a cognitive-based selection instrument. Finally, Fletcher, Lovatt and Baldry (1997) examined the role of TA in a sample of 38 assessment center candidates. A significant, moderate, negative relation between TA and test performance was found (uncorrected $\bar{r} = -.34$).

These studies provide initial evidence that TA might be associated with lower applicant performance on selection instruments. Indeed, it is conceivable that for many jobs TA may introduce variance irrelevant to the predictor construct into the selection test scores, resulting in less accurate prediction of job performance, and a particularly disadvantageous situation for applicants high in TA. Consequently, the second goal of this study was to add to the very limited base of published work that has assessed the relationship between TA and selection test performance by asking the following research question:

Q2: Within personnel selection scenarios, will test anxiety exhibit a negative relation with test performance?

Gender and Test Anxiety

The third goal of the current research was to examine the role of gender in selection TA. A pervasive finding in the TA literature is that higher levels of TA are reported among females than males (Cassady & Johnson, 2002). In fact, Hembree (1988) meta-analyzed the findings of 154 studies of test anxiety and gender, and found strong evidence that females experience higher levels of anxiety than males (mean effect size = .29). This finding, combined with results indicating that TA is negatively related to test performance, has led many to assume that females may be at a disadvantage in testing situations (Zeidner, 1998). In order to properly evaluate this assumption, however, the relationship between TA and test performance must be considered separately for males and females. More specifically, it is essential to consider whether gender moderates the relationship between TA as a predictor, and selection test scores as criteria. This is distinct from overall levels of TA, which simply indicate whether there are mean differences in TA between the male and female groups. If the relation between test performance and TA differs for males and females, the fact that females have higher TA scores does not necessarily imply that females will be unduly disadvantaged by TA (relative to males). In particular, even though females may exhibit higher mean levels of TA than males, if the relationship between TA and test performance is stronger for males, then males may ultimately be more disadvantaged than females in terms of the effect of TA on test scores.

A limited number of investigations, all of which have been conducted in the educational realm, have presented data that provide some insight into the possible moderating role of gender in TA relations. Although many such studies do not allow inferences to be drawn regarding moderator effects (Zeidner, 1998), at least three studies have found

results suggesting that gender does act as a moderator (Sipos, Sipos, & Spielberger, 1986; Smith, Michael, & Hocevar, 1990; Spielberger, 1980). In each case, the relationship between TA and test performance was stronger for males than for females. For example, Smith *et al*. (1990) examined TA among 130 high school students. Results indicated that when exposed to anxiety inducing instructions, highly anxious males demonstrated a greater decrement in performance on a mathematics test than highly anxious females. In another investigation, Sipos *et al*. (1986) administered the TAI to a sample of 120 Hungarian students. Stronger relationships between TA and test performance were found for males ($\bar{r} = -.40$) than for females ($\bar{r} = -.17$). Finally, administering the TAI to several student samples, Spielberger (1980) reported that the relationships between TA and cognitive test scores were stronger for males ($\bar{r} = -.23$) than for females ($\bar{r} = -.15$). Unfortunately, theory-based explanations for the observed differences were not provided in any of these investigations. In addition, although the male-female differences in correlations appear salient, the studies did not statistically test for the possible moderating effect of gender.

Research on the coping styles adopted by men and women provides valuable insight into the aforementioned findings (Tamres, Janicki, & Helgeson, 2002). To be specific, females may be more efficacious in dealing with the autonomic and cognitive sources of stress that they experience during testing situations due to both the *frequency* and the *type* of coping strategies that females typically employ. This would give rise to a weaker negative relation between TA and test performance for females than males. Accordingly, the differential coping responses adopted by male and female applicants may be responsible for the apparent gender differences in TA-test performance relations observed in past research. This forms the basis of our *sex-linked anxiety coping theory*, which is described in detail below.

A considerable amount of research has indicated that female reactions to stress and anxiety are both quantitatively and qualitatively different from those of males. In particular, research suggests that females have a tendency to engage in significantly more coping strategies than their male counterparts. For example, a study by Ryan, Ployhart, Greguras, and Schmit (1998) found that female job applicants were more likely to attend a test preparation program than were males. Females also demonstrate a stronger tendency to regulate stressful emotions, a process that is commonly referred to as Emotion-Oriented coping (Folkman, Lazarus, Dunkel-Schetter, DeLongis, & Gruen, 1986; Long, 1990). Recently, higher levels of coping behaviors among females have been found in a meta-analysis that examined sex differences in a wide range of coping styles (Tamres *et al*., 2002). Females engaged in 16 of 17 coping strategies more often than males, and these strategies included Problem-Oriented coping, Positive Reappraisal of the situation, seeking Social Support, Active Coping, and Positive Self-talk, all of which have obvious relevance to coping with TA. In contrast, males engaged in only one coping strategy, Avoidance-Oriented coping (i.e., a tendency to avoid the situation), significantly more often than females (Tamres *et al*., 2002).

At a superordinate level, coping strategies are commonly categorized as either Problem-Oriented or Emotion-Oriented (Lazarus & Folkman, 1984). Problem-Oriented coping is directed towards altering the source of the threat (e.g., developing an active plan to alter the situation), whereas Emotion-Oriented coping is directed toward altering the emotional response to the threat (e.g., positive self-talk). When males and females are compared, females exhibit higher levels of both Problem- and Emotion-Oriented coping behaviors than males (see Tamres *et al*., 2002). Thus, Tamres *et al*.'s meta-analysis suggests that, on average, males are less likely to engage in coping strategies to deal with TA than are females. Furthermore, the type of coping strategy that males are most likely to adopt (i.e., Avoidance-Oriented) is less likely to be beneficial in testing situations.

In summary, the *sex-linked anxiety coping theory* advanced in this paper suggests that females are more likely to have coping resources available to them to deal with anxiety before selection tests, and are more likely to use adaptive coping behaviors during a selection test. As a result, the relation between TA and test performance is expected to be weaker for female than male test-takers. This theory makes intuitive sense, as females have been found to report higher levels of anxiety across a wide range of situations (Hembree, 1988), which may foster a stronger and more diverse coping repertoire (Long, 1990).

Based on the aforementioned review of past research, as well as our *sex-linked anxiety coping theory*, we offer the following research questions:

Q3a: Will females exhibit significantly higher levels of selection test anxiety than males?

Q3b: Will the relations between selection test anxiety and test performance be moderated by gender, with stronger relations for males than females?

Q3c: Will females exhibit significantly higher levels of Problem-Oriented and Emotion-Oriented coping than males in a selection-testing scenario?

Q3d: Will males exhibit significantly higher levels of Avoidance-Oriented coping than females in a selection testing scenario?

The Current Research

The aforementioned questions were tested in two separate studies. In Study 1, levels of selection TA and scores on a

selection instrument were obtained for a large sample of participants. This provided critical information for Questions 1 and 2, as well as for Questions 3a and b. In Study 2, the coping strategies adopted by males and females when faced with the challenge of a selection test were examined. This provided critical information for Questions 3c and 3d.

Study 1 Method

Participants and Procedure

Participants were 248 undergraduate volunteers (145 females, 103 males) ranging from 17 to 49 years of age ($M = 21.0$, $SD = 4.6$), from a large Canadian university. Participants were informed that their responses were confidential and they were fully debriefed following the study.

Participants were administered a battery of test attitude and cognitive ability items during a two hour group testing session. The number of participants in each session ranged from five to twenty-seven. Upon arrival, participants were asked to complete three forms: informed consent, demographics, and contact information. Following this, the examiner provided instructions for completing the paper and pencil test battery. Participants were advised to pay close attention, as each section of the battery had different instructions and was carefully timed. The test battery consisted of two sections: Section I – test attitudes (including TA) and other attribute measures that were part of a larger study; and Section II – a simulated selection test consisting of a multidimensional measure of cognitive ability. Based on standardized instructions for the respective tests, participants were given 40 minutes to complete the questions in Section I and 36 minutes to complete Section II.

The instructions accompanying the administration of Section I urged the participants to respond to the items with complete honesty. In Section II, participants were asked to play the role of a student applicant who must obtain an impressive score on the tests in order to be selected into a highly desirable University program (e.g., Schmit & Ryan, 1992). Six cash awards were provided for the top scores on the cognitive test in lieu of a true "job" offer and participants were informed of this contingency. It was believed that these cash awards, in lieu of a job offer, would make this testing situation a reasonably accurate reflection of a true selection scenario. Previous use of a similar induction by Schmit and Ryan (1992) proved to be successful. Moreover, an analysis of the mean TA levels exhibited by the current sample indicated that the participants in our study experienced anxiety levels that were comparable to those reported by individuals faced with the challenge of real test-taking situations (Arvey *et al*., 1990; Zeidner, 1991).

Measures

Test Anxiety. Test anxiety was assessed with three distinct measures that were included in the first section of the test battery. The response format for all three scales ranged from 1 (strongly disagree) to 5 (strongly agree). The first scale contained 7 items from the comparative anxiety scale of the Test Attitude Survey (TAS; Arvey *et al*., 1990). The TAS anxiety scale has been found to exhibit internal consistency reliability estimates of .80 (Arvey *et al*., 1990) and .82 (McCarthy & Goffin, 2003). The second anxiety scale was the two-dimensional Test Attitude Inventory (TAI; Spielberger, 1980). The TAI consists of worry and emotionality components (10 items each) and typically has internal consistency reliability of .92 or higher (Spielberger, 1980). The third measure of TA was the 20-item version of Sarason's (1984) four-dimensional Reactions to Tests Scale (RTT). This scale has demonstrated internal consistency reliability of .90 and contains scales measuring tension, bodily symptoms, worry, and test-irrelevant thoughts (Benson & Bandalos, 1992).

Cognitive Ability. A cognitive ability test was included because of the ubiquitous success of this selection testing method (e.g., Hunter, 1986). Three components of cognitive ability were assessed in Section II of the test session: general knowledge, quantitative, and verbal. A multiple-choice format with one correct answer and four distracters was used for all three scales, and scores were based on the percentage of correct responses within each scale. General knowledge and quantitative ability were assessed using the 70-item Personnel Assessment Form (PAF; see Goffin, Gellatly, Paunonen, Jackson, & Meyer, 1996). Verbal ability was measured using 46 items from the vocabulary subscale of the Multidimensional Aptitude Battery (MAB; Jackson, 1985). Each of these scales has been found to demonstrate adequate psychometric properties (Goffin *et al*., 1996; Jackson, 1985). In addition to examining the relations involving general knowledge, quantitative ability, and verbal ability separately, a cognitive ability composite was formed by unit-weighting and adding the three cognitive scales.

Grade Point Average. Participants' grade point average (GPA) was calculated as the average score across the five most recently completed courses, and ranged from 0 to 100.

Infrequency. In order to detect non-purposeful or random responding, the 16-item infrequency scale of the Personality Research Form (Jackson, 1987) was administered to participants in the first section of the test battery. Jackson (1987) reported an internal consistency reliability of .71 for this scale using a college sample.

Analytic Strategy

With respect to the first research question (Q1), the dimensionality of TA was assessed using confirmatory factor analyses. Three structures were examined: the

unidimensional TA structure presumed by Mandler & Sarason (1952), the two-dimensional structure (i.e., worry and emotionality) presumed by Liebert & Morris (1967), and the four-dimensional structure (i.e., tension, bodily symptoms, worry and test-irrelevant thoughts) presumed by Sarason (1984). Once the most acceptable TA dimensional structure was confirmed through the factor analysis, it was possible to provide basic psychometric information on the derived TA subscales that corresponded to the factors. The second research question (Q2: the relation between TA and test performance) was then examined by conducting a series of correlational analyses. Finally, the third research question (Q3: gender differences in TA and the role of gender as a moderator) was explored using a series of statistical tests and moderated multiple regression analyses.

Study 1 Results

Preliminary Analyses

Twelve participants were eliminated due to high infrequency scores (i.e., average item scores higher than 2.00 on the infrequency scale). This resulted in a final sample of 242 participants (142 females).

Dimensionality of Test Anxiety

Confirmatory factor analyses were conducted on the variance–covariance matrices using Amos 4.10 (Arbuckle, 1999). Maximum likelihood estimation was used, and four indices were employed to assess the fit of the models: the Chi-Square index, the Root Mean Square Error of Approximation (RMSEA, Steiger, 1989), the Relative Noncentrality Index (RNI, Goffin, 1993; McDonald & Marsh, 1990), and the Normed Fit Index (NFI; Bentler & Bonnett, 1980). This combination of fit indices ensured the inclusion of an index that considers how much variance is explained in light of how many degrees of freedom are used (i.e., RMSEA), as well as an index that is a direct function of how much variance is explained by the model (i.e., RNI; Goffin, 1993). For the RMSEA, MacCallum, Browne, and Sugawara (1996) have suggested that values in the .08–.10 range indicate a mediocre fit, values in the .05–.08 range indicate a fair fit, and values less than .05 indicate a close fit. In the case of the RNI and NFI, values approaching 1.0 indicate good fit. Factors were allowed to correlate, as the dimensions of TA were expected to be related. Items were not allowed to load on more than one factor.

The one-factor model consisted of all 47 items from the TAS, TAI and RTT test anxiety scales, the two-factor model consisted of 38 items because nine of the items in the one-factor model could not reasonably be assigned to either the emotionality or worry factors, and the four-factor model contained 44 items because three of the items in the one-factor model could not reasonably be assigned to either the tension, bodily symptoms, worry, or test-irrelevant thoughts factors. Although 38 of the items were common across all models, nested model comparisons were not appropriate because the models did not involve an identical set of items. Therefore, the assessment of relative fit was based on the RMSEA and RNI values. The one-factor model ($\chi^2 = 2779.19$, df = 1034) yielded an RMSEA of .084, an RNI of .94, and an NFI of .91. The two-factor model ($\chi^2 = 1435.90$, df = 664) yielded an RMSEA of .069, an RNI of .97, and an NFI of .93. The intercorrelation of the two factors was .83. The four-factor model ($\chi^2 = 1971.89$, df = 896) yielded an RMSEA of .071, an RNI of .96, and an NFI of .93. The intercorrelations of the four factors ranged from .37 to .91.

Overall, the two-factor model demonstrated the best fit to the data. Although the χ^2 was significant (as is typically the case even when the fit is acceptable; McDonald & Marsh, 1990), the RNI was highest for the two-factor model (RNI = .97). In addition, the RMSEA for this model (.07) was acceptable. Moreover, with the exception of two items, all standardized loadings for the two-factor model were greater than .30 (average = .65), and were significantly different from zero. A series of hierarchical regression analyses also supported the bi-dimensional structure. To be specific, we assessed the extent to which factor 1 (emotionality) adds incremental variance in test performance over and beyond factor 2 (worry), and vice versa. Findings revealed that although emotionality did not add to the predictive validity in overall cognitive ability above and beyond worry ($\Delta R^2 = .01$, $F = 2.37$, n.s.), worry did indeed add incremental explanatory variance in overall cognitive ability above and beyond emotionality ($\Delta R^2 = .03$, $F = 9.15$, $p < .01$). These results demonstrate the discriminant validity of the two-factor model.

Finally, hierarchical regression analyses were conducted to examine whether the four-factor model added to the prediction of performance beyond the two-factor model. This analysis was particularly meaningful, as the two-factor model consists of emotionality and worry components, while the four-factor model consists of emotionality, worry, test-irrelevant thoughts and bodily symptom components. Therefore, we assessed whether the two additional factors that were identified in the four-factor model of TA (test-irrelevant thoughts and bodily symptoms) added to the prediction of test performance beyond the factors that corresponded to the two-factor model of TA (emotionality and worry). Results indicated that the test-irrelevant thoughts and bodily symptoms factors did *not* add to the prediction of cognitive test performance beyond the worry and emotionality factors ($\Delta R^2 = .00$, $F = 0.73$, n.s.). In contrast, the emotionality and worry factors did add to the prediction of performance beyond the test-irrelevant thoughts and bodily symptoms factors ($\Delta R^2 = .05$, $F = 7.42$, $p < .001$). Combined, these analyses provide further support for the superiority of the two-factor model.

On the basis of the aforementioned analyses, TA-Emotionality (i.e., tension) and TA-Worry (i.e., fear of failure) scales were created by unit-weighting and aggregating the respective items.

Psychometric Properties of the TA Scales

Table 1 presents the descriptive statistics and internal consistency reliabilities of study variables. As indicated, the internal consistency reliabilities of the TA and cognitive ability variables were high (.83–.96). Also presented in Table 1 is the relation between the two TA scales. As illustrated, this relation is positive and relatively strong ($r = .73$).

Relations Between Test Anxiety and Test Performance

Table 1 presents the correlations between TA and the selection tests. As illustrated, the pattern of relations was consistent for both TA scales – TA was negatively related to performance on all three cognitive ability scales, as well as the overall cognitive ability composite scale. One exception was a non-significant relationship found between TA-Worry and quantitative ability.

Hierarchical regression analyses were also conducted to assess the magnitude of the relation between TA and overall performance on the cognitive ability test with the effect of prior ability (indexed as grade point average) partialled out of the equation. This enabled an indirect assessment of whether poor test performance is due to poor ability, causing high levels of TA, which in turn causes lower test performance. Findings indicated that the TA-Emotionality and TA-Worry factors added a significant amount of incremental validity to the prediction of overall cognitive test performance above and beyond the prediction obtained by prior ability ($\Delta R^2 = .11$, $p < .01$). This finding is equivalent to a multiple correlation of .33, and suggests that TA has an effect on test performance that is independent of prior ability levels.

The Role of Gender

The third goal of the present investigation was to examine potential male–female differences in the relationships of the two TA scales with the selection test. Prior to examining male–female differences in relations, it was of interest to determine whether male–female mean differences existed in any of the measures. Consistent with research question 3a, significant mean differences were obtained for both the TA-Emotionality ($t_{(240)} = 2.46$, $p < .05$), and TA-Worry ($t_{(240)} = 2.09$, $p < .05$) scales (see Table 2). In both cases, females exhibited higher scores.

To examine research question 3b, which asked whether gender moderates the relationship between TA and scores on the selection instruments, moderated multiple regressions were performed (see Table 3). This is analogous to testing for slope differences across gender in the

Table 1. Means, standard deviations, internal consistency reliabilities, and intercorrelations of study variables

Variables	*M*	SD	1	2	3	4	5	6	7	8	9	10
Study 1												
1. TA-Emotionality	2.77	.80	*.96*									
2. TA-Worry	2.40	.63	.73**	*.89*								
3. CA general knowledge	2.92	.78	−.31**	−.32**	*.83*							
4. CA quantitative	1.72	.49	−.15*	−.11	.36**	[1]						
5. CA verbal	2.95	.93	−.27**	−.35**	.59**	.14*	*.83*					
6. CA composite	2.53	.57	−.33**	−.37**	.87**	.53**	.85**	[1]				
7. GPA	80.00	6.08	−.05	−.32**	.19**	.24**	.29**	.31**	[2]			
Study 2												
8. Problem-Oriented coping	3.70	.49	–	–	–	–	–	–	–	*.65*		
9. Emotion-Oriented coping	3.37	.50	–	–	–	–	–	–	–	.25**	*.61*	
10. Avoidance-Oriented coping	2.70	.69	–	–	–	–	–	–	–	−.33**	−.13	*.64*

Notes: In study 1, *N* ranged from 191 to 242. In study 2, $N = 160$. TA, Test Anxiety; CA, Cognitive Ability Test. With the exception of GPA, item scores ranged from 1 to 5. GPA scores ranged from 0 to 100. Internal consistency reliabilities are on the diagonal.

[1] α could not be computed for the respective measure either because it is a speeded test (i.e., Quantitative) or contains a speeded test (i.e., CA Composite), and α is not an appropriate measure of internal consistency for speeded tests (Anastasi, 1982).

[2] α is not relevant as GPA is a single-item scale.

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

Table 2. Gender differences in test anxiety, cognitive ability, and personality

Measures	Females		Males		*t*-test of means	*F*-test of standard deviations	*F*-test of reliabilities
	M (SD)	α	*M* (SD)	α			
Test anxiety							
TA-Emotionality	2.88 (.83)	.97	2.62 (.74)	.95	$t_{(240)} = -2.46$, $p < .05$	[2]	[3]
TA-Worry	2.47 (.65)	.90	2.30 (.59)	.87	$t_{(240)} = -2.09$, $p < .05$	[2]	[3]
Cognitive ability							
General knowledge	2.88 (.78)	.83	2.96 (.78)	.84	$t_{(238)} = -0.74$, n.s.	$F_{(139)} = 1.00$, n.s.	$F_{(139,99)} = 1.06$, n.s.
Quantitative	1.69 (.45)	[1]	1.76 (.53)	[1]	$t_{(238)} = -1.00$, n.s.	$F_{(139)} = 1.18$, n.s.	[1]
Verbal	2.92 (.90)	.82	2.99 (.97)	.84	$t_{(238)} = -0.59$, n.s.	$F_{(139)} = 1.08$, n.s.	$F_{(139,99)} = 1.13$, n.s.
CA composite	2.50 (.57)	[1]	2.57 (.58)	[2]	$t_{(238)} = -0.94$, n.s.	$F_{(139)} = 1.02$, n.s.	[1]

Notes: $N = 242$ (142 females, 100 males). CA Composite, Cognitive Ability Composite; AD Composite, Achievement/Dominance Composite. Cognitive ability scores were scaled to range from 1 to 5. α, alpha reliability.
[1]α could not be computed for the quantitative subscale because it is a speeded test, and α is not an appropriate measure of internal consistency for speeded measures (Anastasi, 1982).
[2]The *F*-test was not conducted because the variance of female scores was higher that that of males.
[3]The *F*-test was not conducted because the reliabilities of female's scores were higher than that of males.

Table 3. Gender as a moderator of the relation between test anxiety and test performance

Relation	*r* (females)	*r* (males)	Test of slope differences	Test of intercept differences
TA-Emotionality and . . .				
Cognitive Ability				
General Knowledge	−.24**	−.42**	$\Delta R^2 = .01$, $F = 2.77$, n.s.	$\Delta R^2 = .00$, $F = 0.00$, n.s.
Quantitative	−.06	−.27**	$\Delta R^2 = .02$, $F = 4.02$, $p < .05$	$\Delta R^2 = .02$, $F = 4.43$, $p < .05$
Verbal	−.24**	−.33**	$\Delta R^2 = .00$, $F = 1.40$, n.s.	$\Delta R^2 = .00$, $F = 0.00$, n.s.
CA Composite	−.25**	−.45**	$\Delta R^2 = .02$, $F = 4.05$, $p < .05$	$\Delta R^2 = .01$, $F = 3.78$, n.s.
TA-Worry and . . .				
Cognitive Ability				
General knowledge	−.19*	−.52**	$\Delta R^2 = .03$, $F = 7.85$, $p < .05$	$\Delta R^2 = .03$, $F = 7.63$, $p < .05$
Quantitative	−.02	−.21*	$\Delta R^2 = .01$, $F = 2.92$, n.s.	$\Delta R^2 = .01$, $F = 0.48$, n.s.
Verbal	−.26**	−.48**	$\Delta R^2 = .02$, $F = 6.50$, $p < .05$	$\Delta R^2 = .02$, $F = 5.82$, $p < .05$
CA Composite	−.23**	−.57**	$\Delta R^2 = .04$, $F = 10.37$, $p < .05$	$\Delta R^2 = .04$, $F = 9.97$, $p < .05$

Note: $N = 242$ (142 females, 100 males). CA Composite, Cognitive Ability Composite; AD, Achievment/Dominance Composite.

relation between TA and test scores (Lautenschlager & Mendoza, 1986). The relation between TA-Emotionality and test performance was considered first. As indicated in the table, gender significantly moderated the relationship between TA-Emotionality and the Cognitive Ability Composite ($\Delta R^2 = .02$, $F = 4.05$, $p < .05$). However, at the subscale level only quantitative ability ($\Delta R^2 = .02$, $F = 4.02$, $p < .05$) was significantly moderated by gender. The relations between TA-Worry and test performance were considered next. Findings revealed that gender significantly moderated the relationship between TA-Worry and the Cognitive Ability Composite ($\Delta R^2 = .04$, $F = 10.37$, $p < .01$). At the subscale level, the relations involving general knowledge ($\Delta R^2 = .03$, $F = 7.85$, $p < .01$) and verbal ability ($\Delta R^2 = .02$, $F = 6.50$, $p < .05$) were significantly moderated by gender. Overall, these findings suggest that the negative slope between TA and performance on selection tests is stronger for males.

In addition to testing for differences in slope, we examined potential intercept differences using the procedure recommended by Lautenschlager and Mendoza (1986). Findings are presented in the final column of Table 3. As illustrated, significant slope and intercept differences

were found for the relations between: TA-Emotionality and quantitative ability; TA-Worry and general knowledge; TA-Worry and verbal ability; and TA-Worry and the Cognitive Ability Composite. Follow-up analyses revealed that in each case, if a female reports a low level of anxiety, then her test score will be overpredicted by the regression line that is based on the entire sample. In contrast, if a female reports a high level of anxiety, then her test score will be underpredicted by the regression line that is based the entire sample. The exact opposite trend occurred for males.

Exploration of Statistical Artifacts

As reported, gender significantly moderated the majority of the relations of TA-Emotionality and TA-Worry with selection test scores. For the sake of thoroughness, in this section we examined some possible artifactual explanations for the observed moderator effects.

First, gender moderation, in the form of lower TA-test performance relations for females, could occur as a consequence of ceiling effects that result in range restriction. These ceiling effects and the resulting restriction of range could occur due to the higher levels of TA that are typically observed in females. This possibility was assessed by examining whether the variance in TA was significantly greater for males than for females (see the fifth column of Table 2). As indicated in Table 2, this explanation was not supported – findings indicated that the standard deviations of the TA-Emotionality and TA-Worry scales were slightly higher for females. Visual inspection of the graphs representing male and female TA scores also suggested that ceiling effects were not evident.

Second, gender moderation may be caused by a restriction of range in females' scores on the selection test. This would have the effect of attenuating the correlation between females' TA and test performance. This was explored by examining whether the variance in cognitive ability scores was significantly greater in males than in females (see the fifth column of Table 2). Results were unequivocal – in all cases the standard deviation of males' scores was not significantly higher than the standard deviation of females'. In addition, visual inspection of the graphs representing male and female cognitive ability scores suggested that floor and/or ceiling effects were not evident for either gender. Accordingly, the second possible artifactual explanation for gender moderation was dismissed.

A third possible statistical explanation regarding the higher male than female relations of TA scores with cognitive ability test scores is that females' scores on the predictors and/or criteria were less reliable than males'. This was explored by statistically testing whether or not males' alpha reliabilities of TA and cognitive ability were higher than females' using the *F*-test discussed in Feldt, Woodruff, and Salih (1987). As shown in the final column of Table 2, males' general knowledge and verbal alphas were slightly higher than females' but the differences were not significant. In sum, it appears unlikely that artifacts related to restriction of range or reliability could explain the gender moderation effect.

Finally, the possibility of gender bias in the cognitive ability tests was assessed by conducting moderated multiple regression analyses with TA as the predictor, GPA as the criterion, and gender as the moderator. These analyses were conducted separately for the three cognitive ability scales (General Knowledge, Quantitative, and Verbal), and for the two dimensions of TA (Emotionality and Worry). In all cases the slope of the regression line between cognitive ability test scores and GPA did *not* differ depending on whether participants were male or female. However, significant intercept differences were found as a result of females exhibiting higher overall levels of GPA than males. The finding that females in this age range have higher GPAs than males is consistent with past research (see Hyde & Kling, 2001).

Study 2 Method

Participants, Procedure, and Measures

Given that gender moderation could not be explained by statistical artifacts, an exploration of the *sex-linked anxiety coping theory* put forth in the introduction was conducted. This was accomplished by having 164 undergraduate management students (92 females, mean age = 20.5) play the role of a job applicant who is highly motivated to obtain an impressive score on a selection test in order to be selected into a prestigious company. They were than asked to complete the Ways of Coping Questionnaire (Folkman & Lazarus, 1988) in order to assess what coping strategies they would most likely use in dealing with the upcoming selection test. The coping items were organized into the three relevant categories discussed in our earlier section on the sex-linked anxiety coping theory and described in the Tamres *et al*. (2002) meta-analysis: Problem-Oriented Coping (i.e., engaging in behaviours to alter the stressor; eight items, $\alpha = .65$), Emotion-Oriented Coping (i.e., seeking comfort or emotional support from others; five items, $\alpha = .61$), and Avoidance-Oriented Coping (i.e., engaging in behavioural and mental efforts to keep one's mind off the stressor; five items, $\alpha = .64$). Responses were made on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5).

Study 2 Results

Preliminary Analyses

Four participants were eliminated due to high infrequency scores (i.e., average item scores higher than 10.00 on the infrequency scale – had four items 1–5). This resulted in a final sample of 160 participants (90 females).

Table 4. Gender differences in coping with test anxiety

Coping variable	Females *M* (SD)	Males *M* (SD)	*t*-test
Problem-Oriented	3.77 (.47)	3.61 (.51)	$t_{(158)} = -2.00, p < .05$
Emotion-Oriented	3.51 (.52)	3.19 (.41)	$t_{(158)} = -4.16, p < .05$
Avoidance-Oriented	2.56 (.66)	2.87 (.71)	$t_{(158)} = 2.90, p < .05$

Note: $N = 160$ (90 females, 70 males).

Gender Differences in Coping Strategies

A series of *t*-tests were conducted in order to examine the *sex-linked anxiety coping theory* (see Table 4). Females exhibited significantly higher levels of Problem-Oriented and Emotion-Oriented coping strategies than their male counterparts. In contrast, males exhibited higher levels of Avoidance-Oriented coping than their female counterparts. Finally, as a group, males exhibited higher levels of Problem-Oriented ($M = 3.61$, SD = .51) than Emotion-Oriented ($M = 3.19$, SD = .41) coping ($t_{(69)} = 59.10$, $p < .001$). Collectively, these findings provide support for the *sex-linked anxiety coping theory.*

Discussion

This study was designed to assess (a) the dimensionality of TA in a personnel selection context, (b) the relationship between TA and performance on a selection instrument, and (c) the role of gender in selection TA. Although TA has been studied in the educational realm, the construct of TA has received insufficient consideration with respect to selection testing. This is surprising, as the extreme evaluative and competitive stressors experienced in the selection testing process (Rynes, Bretz, & Gerhart, 1991) do not appear to be readily interchangeable with those experienced in educational testing. Moreover, although the published literature provides some hints that gender may moderate the relations of TA with test performance in the educational realm (Sipos, Sipos, & Spielberger, 1986; Smith, Michael, & Hocevar, 1990; Spielberger, 1980), gender moderation of TA-test performance relations has not been systematically evaluated in educational or selection testing contexts. Accordingly, the current study examined the relationship between TA, gender, and test performance in a selection-testing scenario. To our knowledge, this is also the first investigation that has explicitly tested for gender moderation in the relation between TA and test performance. The current study also breaks new ground by advancing and testing the *sex-linked anxiety coping theory.* Consequently, this research represents a significant contribution to the field, and, as discussed in the proceeding sections, the findings provide valuable implications for job applicants and organizations, as well as several important directions for future research.

The Dimensionality of Test Anxiety

The theory-driven approach that was used in this study to investigate the dimensionality of TA served to advance past research, as it represented the first confirmatory analysis of items taken from a diverse set of TA instruments. It was also the first study to investigate the dimensionality of TA within a personnel selection scenario, thus providing unique insight into TA with respect to selection testing contexts.

Findings supported a bi-dimensional structure of TA that consists of emotionality (i.e., tension) and worry (i.e., fear of failure) components. The factors directly reflect Liebert and Morris' (1967) two-component model of TA, suggesting that TA has both an autonomic (emotionality) and a cognitive (worry) component in selection contexts. However, the .73 correlation between the two scales (see Table 1) suggests that the construct of TA may be hierarchical in nature, consisting of a superordinate general TA anxiety factor and two underlying dimensions. Taxonomically, this suggests that the higher order factor underlying emotionality and worry is directly related to anxiety, which itself is a facet of the neuroticism factor of the Big 5 model of personality (Costa & McCrae, 1992).

Relation Between Test Anxiety and Test Performance

The second goal of this study was to add to the meager literature base that has assessed the relation between TA and selection test performance. Results indicated that both TA-Emotionality and TA-Worry were negatively related to performance on the Cognitive Ability Composite ($\bar{r} = -.35$). This finding is consistent with research examining the relationship between TA and academic test performance ($\bar{r} = -.18$; Hembree, 1988). It is noteworthy, however, that the relationship was considerably stronger in the present investigation. In fact, our results parallel the findings obtained by Arvey *et al.* (1990), which demonstrated a moderate negative correlation between TA and scores on a cognitive ability measure in a selection context ($\bar{r} = -.35$). Current findings are also consistent with research in the realm of job interview anxiety, where a multiple correlation of $-.34$ between interview anxiety

and interview performance has been found (McCarthy & Goffin, 2004).

Although the aforementioned findings suggest that the construct of TA is negatively related to performance on an ability test, it is important to note that data for the current investigation were correlational in nature. Consequently, it was not possible to ascertain whether high levels of anxiety resulted in lower levels of test performance, or whether low levels of test performance prompted participants to report high levels of anxiety. However, although a direct assessment of causality could not be conducted, it was possible to assess the magnitude of the relation between TA and performance on the ability test with the effect of prior ability (indexed as grade point average) partialled out of the equation. This enabled an indirect assessment of whether poor test performance is due to poor ability, causing high levels of TA, which in turn causes lower performance. Findings indicated that the TA-Emotionality and TA-Worry factors added a significant amount of incremental validity to the prediction of cognitive test performance above and beyond the prediction obtained by prior ability. This is consistent with research by Chan, Schmitt, DeShon, Clause, and Delbridge (1997), who found that test-taking attitudes were related to later test performance when the effects of performance on the first test were controlled. Nevertheless, additional research, possibly using time series and/or experimental designs, is required to explore the directionality of the relation between TA and ability.

Finally, the pattern of relationships between TA and the three cognitive ability subscales (general knowledge, quantitative and verbal) was examined. A differential pattern of relations was found, in that the quantitative scale did not correlate nearly as highly with the TA-Worry and TA-Emotionality scales as did the verbal and general knowledge scales. One implication of this finding is that scores on quantitative ability may be more "hard-wired" and not as easily improved by lessening a candidate's levels of TA. Future research is required to explore this possibility.

The Role of Gender

The third goal pursued in this investigation was to investigate the role of gender in selection TA. Consistent with previous findings, females exhibited higher mean levels of TA-Emotionality and TA-Worry than males (Seipp, 1991; Zeidner, 1998). In addition, gender significantly moderated the relations between (a) TA-Emotionality and performance the quantitative ability and composite scale, and (b) TA-Worry and performance on the general knowledge, verbal ability, and composite scale. In all cases, the slope of the relationship between TA and test performance was *stronger for males*. This is consistent with the *sex-linked anxiety coping theory* that is advanced in the current paper. In addition, significant intercept differences were noted in four of the eight moderated multiple regression analyses. These results are consistent with the well-known finding that females exhibit higher levels of TA than males (Seipp, 1991).

An examination of three possible artifactual explanations for the moderator effect were examined. Findings suggested that the gender moderation was *not* solely the result of higher variance and/or higher reliability of males' scores. It therefore appears that the gender moderation was *not* due to statistical artifacts, and should be considered from a theoretical lens. The theoretical possibility that we examined is the *sex-linked anxiety coping theory*. As previously described, it is possible that females may be more efficacious at coping with anxiety in selection testing situations because they are more likely to have coping resources available to them before the test, and are more likely to use adaptive coping behaviors during the test. This may have the effect of reducing the impact of TA on subsequent test performance. Study 2 was designed to explore this possibility and found that females exhibited significantly higher levels of Problem-Oriented and Emotion-Oriented coping than males (Q3c), and males exhibited significantly higher levels of Avoidance-Oriented coping than females (Q3d). Findings also revealed that as a group, males exhibited significantly higher levels of Problem-Oriented than Emotion-Oriented coping. These findings provide initial support for our *sex-linked anxiety coping theory* and provide a foundation for future work assessing gender differences in selection TA, as well as other types of TA (i.e., educational).

Implications and Conclusions

We acknowledge that there are limits to the generalizability of the current work owing to the student sample and the lab-based design. However, students are exposed to many tests at the elementary, secondary, and undergraduate levels of education (Ross, 1999). The fact that TA effects were found with them suggests that the effects may be stronger with traditional job applicants, who may have been exposed to fewer tests. As a result, the current findings may underestimate the true relations.

In addition, as previously discussed, TA can be conceptualized as a situation-specific trait that is manifested in *any* testing situation, including personnel selection contexts. Indeed, considerable effort was expended to make the selection simulation as realistic as possible in order to engender a level of TA that would approximate the level experienced by true job applicants writing a selection test. For example, the "job" being applied for was one that students were very familiar with and experienced at. In addition, cash rewards were offered to top scorers on the selection tests, an approach that has been successful in the past in simulating the demand characteristics of selection situations (Schmit & Ryan, 1992). Nevertheless,

replication of this work in a field setting would be a worthwhile endeavor for future research.

Several particularly meaningful implications can be derived from the findings of this investigation. First, it appears likely that TA assumes a bi-dimensional structure, worry and emotionality in personnel selection contexts. Second, the negative relations between test performance and the TA-Worry and TA-Emotionality scales suggest that TA may introduce irrelevant variance into selection test scores, resulting in less accurate prediction of job performance and a particularly disadvantageous situation for applicants high in TA. For example, high levels of TA may result in low-test scores, in spite of the fact that candidates may demonstrate superior on-the-job performance if hired. In support of this proposition, Schmit and Ryan (1992) found that the predictive validity of a selection instrument was *lower* for individuals with high levels of anxiety. It is important to acknowledge that this conceptualization of TA assumes that TA has little bearing on subsequent job performance. Admittedly, maintaining composure in high stress situations may be an important competency for certain jobs (i.e., air traffic controller, surgeon). However, the high-stakes selection-testing situation that is experienced by many job applicants is relatively uncommon in the daily activities of most employees, including, for example, administrative personnel, mid-level managers, and service workers. Indeed, the average job does not require employees to write highly evaluative tests on a regular basis. In fact, a recent American Medical Association Survey on workplace testing (1999) found that 83% of job incumbents are not required to write ability tests. Moreover, most jobs promotions are more likely to be based on the demonstration of skills than on the writing of tests. Therefore, the successful handling of test anxiety is *unlikely* to be a valid measure of performance for most jobs, and TA may serve to bias the predictive validity of selection tests. Future research that directly examines whether TA obscures the prediction of applicants' subsequent job performance, and ultimately results in the selection of less promising candidates would be valuable.

The observed negative relations between TA and test performance also attest to the potential importance of reducing TA among job applicants. Indeed, it is in the best interest of both organizations and employees to find ways to reduce TA so that a more accurate picture of applicant skills can be obtained. In this regard, there is evidence that providing individuals with choice in test items enhances their perceived feelings of control, which decreases their anxiety (Keinan & Zeidner, 1987). Computerized tests are one means through which examinees might be offered some degree of choice over which items they respond to (Zeidner, 1998). Moreover, there is evidence that computerized testing may result in lower levels of anxiety than paper and pencil tests because it can provide a less competitive testing environment due to individualized, instead of group, administration procedures (Zeidner, 1998). Lower levels of anxiety have also been reported for untimed tests (Plass & Hill, 1986). In fact, a study investigating student reactions to the Graduate Record Exam found that time restrictions were the primary cause of student anxiety (Powers, 1986). Regardless, considerable ingenuity may be required in order to modify selection testing in such a way as to reduce the TA it induces while not compromising the vital purposes it serves.

The current gender moderation results imply that, contrary to popular belief, females may *not* be at a particular disadvantage in testing situations. Although females exhibited higher levels of TA on average, and a negative relationship between TA and test performance was found, the relationship between TA and test performance was actually found to be higher for male applicants. As illustrated in Study 2, this appears to be a result of differential coping strategies that are employed by male and female applicants when faced with an upcoming selection test. To be specific, consistent with the *sex-linked coping theory*, females are more likely to use positive coping strategies (i.e., Problem-Oriented and Emotion-Oriented coping) than males. The implication of this finding is substantial, as research indicates that both women and men can be trained in both Problem-Oriented and Emotion-Oriented coping strategies (Bekker, Nijssen, & Hens, 2001). Thus, future research assessing the extent to which training in coping strategies can result in the reduction of TA and improvement of test scores for male and female applicants would be particularly advantageous.

It is also important to note that the stronger relation between TA and test performance for males implies that TA may more strongly bias the test performance of males, which may ultimately result in lower predictive validity estimates for male applicants. Existing research lends support to this possibility, as four studies have found that the relationship between ability test scores and job performance are lower for males (Rothstein & McDaniel, 1992; Schat & Hausdorf, 2000; Schmitt, Mellon, & Bylenga, 1978; Swarthout *et al*., 1984, as cited in Hartigan & Wigdor, 1989). Nonetheless, the fact that females experienced higher mean levels of TA in this work and earlier work (Hembree, 1988) is still cause for concern as it suggests that the subjective well-being of females is likely to be notably lower than that of males in selection testing contexts. Techniques to reduce TA, such as relaxation training, for both male and female applicants are therefore required. In addition, future research on reducing the potentially harmful effects of TA should be sensitive to the possibility of gender differences and gender moderation. Additionally, the stereotype threat literature (i.e., Oswald & Harvey, 2000; Steele, 1997; Walsh, Hickey, & Duffy, 1999) may provide clues as to the etiology of gender differences in TTA. In particular, future research that measures distinct levels of applicant TA with respect to quantitative and verbal ability tests would enable a

thorough assessment of the role of stereotype threat and TA.

In conclusion, the present investigation provided valuable insight on TA in the context of personnel selection. In doing so, it answers a recent call for emphasis on measurement issues when studying applicant reactions to selection procedures (Ryan & Ployhart, 2000), and is consistent with recommendations to examine new applicant perceptions, such as TA (Ryan & Ployhart, 2000). As indicated, there are several important avenues for future research and it is our hope that this study will foster additional work in this exciting new realm.

Acknowledgements

We are grateful to Ryan Stanga, Andre Popadynec, and Henryk Krajewski for their help with this project. This work was supported by separate grants from the Social Sciences Research Council of Canada to Julie McCarthy and Richard Goffin.

References

American Medical Association. (1999) AMA survey on workplace testing. *Management Review*, **7**, 44–48.

Anastasi, A. (1982) *Psychological testing* (5th Edn.). New York: Macmillan.

Arbuckle, J.L. (1999) *Amos 4.01*. Chicago, IL: Smallwaters Corporation.

Arvey, R.D., Strickland, W., Drauden, G. and Martin, C. (1990) Motivational components of test-taking. *Personnel Psychology*, **43**, 695–717.

Baron, R.M. and Kenny, D.A. (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, **51**, 1173–1182.

Bekker, M.H.J., Nijssen, A. and Hens, G. (2001) Stress prevention training: Sex differences in types of stressors, coping, and training effects. *Stress and Health*, **17**, 207–218.

Benson, J. and Bandalos, D. (1992) Second-order confirmatory factor analysis of the Reactions to Tests Scale with cross validation. *Multivariate Behavioral Research*, **27**, 459–487.

Bentler, P.M. and Bonnett, D.G. (1980) Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, **88**, 588–606.

Cassady, J.C. and Johnson, R.E. (2002) Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, **27**, 270–295.

Chan, D., Schmitt, N., DeShon, R.P., Clause, C.S. and Delbridge, K. (1997) Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, **82**, 300–310.

Cleary, T.A. (1968) Test bias: Predication of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, **5**, 115–124.

Costa, P.T. and McCrae, R.R. (1992) *NEO personality inventory – revised: manual*. Lutz, FL: Psychological Assessment Resources.

Dobson, P. (2000) An investigation into the relationship between Neuroticism, Extraversion and cognitive test performance in selection. *International Journal of Selection and Assessment*, **8**, 99–109.

Elliot, A.J. and McGregor, H.A. (1999) Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, **76**, 628–644.

Endler, N.S., Edwards, J.M. and Vitelli, R. (1991) *Endler multidimensional anxiety scales: Manual*. Los Angeles, CA: Western Psychological Services.

Feldt, L.S., Woodruff, D.J. and Salih, F.A. (1987) Statistical inference for coefficient alpha. *Applied Psychological Measurement*, **11**, 93–103.

Fletcher, C., Lovatt, C. and Baldry, C. (1997) A study of trait, and test anxiety, and their relationship to assessment center performance. *Journal of Social Behavior and Personality*, **12**, 205–214.

Folkman, S. and Lazarus, R.S. (1988) *Manual for the Ways of Coping Questionnaire*. Palo Alto, CA: Consulting Psychologists Press.

Folkman, S., Lazarus, R.S., Dunkel-Schetter, C., DeLongis, A. and Gruen, R.J. (1986) Dynamics of a stressful encounter: Cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, **50**, 992–1003.

Goffin, R.D. (1993) A comparison of two new indices for the assessment of fit of structural equation models. *Multivariate Behavioral Research*, **28**, 205–214.

Goffin, R.D., Gellatly, I.R., Paunonen, S.V., Jackson, D.N. and Meyer, J.P. (1996) Criterion validation of two approaches to performance appraisal: The Behavioral Observation Scale and the Relative Percentile Method. *Journal of Business and Psychology*, **11**, 23–33.

Goffin, R.D., Rothstein, M.G. and Johnston, N.G. (2000) Personality and job performance: Are personality tests created equal? In R.D. Goffin and E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 249–264). Norwell, MA: Kluwer Academic Publishers.

Hartigan, J.A. and Wigdor, A.K. (1989) *Fairness in employment testing: Validity generalization, minority issues and the general aptitude test battery*. Washington, DC: National Academy Press.

Hembree, R. (1988) Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, **58**, 7–77.

Hunter, J.E. (1986) Cognitive ability, cognitive aptitudes, job knowledge and job performance. *Journal of Vocational Behavior*, **229**, 340–362.

Hyde, J.S. and Kling, K.C. (2001) Women, motivation and achievement. *Psychology of Women Quarterly*, **25**, 364–378.

Jackson, D.N. (1965) *Personality research form-BB*. Port Huron, MI: Sigma Assessment Systems.

Jackson, D.N. (1985) *Multidimensional aptitude battery*. Port Huron, MI: Sigma Assessment Systems.

Jackson, D.N. (1987) *Personality research form-E*. Port Huron, MI: Sigma Assessment Systems.

Kalechstein, P., Hocevar, D., Zimmer, J.W. and Kalechstein, M. (1989) Procrastination over test preparation and test anxiety. In R. Schwarzer, H.M. van der Ploeg and C.D. Spielberger *Advances in Test Anxiety Research* (Vol. 6, pp. 63–76). Lisse: Swets & Zeitlinger.

Keinan, G. and Zeidner, M. (1987) Effects of decisional control on test anxiety and achievement. *Personality and Individual Differences*, **8**, 973–975.

Kuncel, N.R., Hezlett, S.A. and Ones, D.S. (2001) A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, **127**, 162–181.

Lautenschlager, G.J. and Mendoza, J.L. (1986) A step down hierarchical multiple regression analysis for examining

hypotheses about test bias in prediction. *Applied Psychological Measurement*, **10**, 133–139.
Lazarus, R.S. and Folkman, S. (1984) *Stress, appraisal, and coping*. New York: Springer.
Liebert, R.M. and Morris, L.W. (1967) Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, **20**, 975–978.
Long, B.C. (1990) Relation between coping strategies, sex-typed traits, and environmental characteristics: A comparison of male and female managers. *Journal of Counseling Psychology*, **37**, 185–194.
MacCallum, R.C., Browne, M.W. and Sugarawa, H.M. (1996) Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, **1**, 130–149.
Mandler, G. and Sarason, S. (1952) A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, **47**, 166–173.
McCarthy, J.M. and Goffin, R.D. (2003) Is the test attitude survey psychometrically sound? *Educational and Psychological Measurement*, **63**, 446–464.
McCarthy, J.M. and Goffin, R.D. (2004) Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology*, **57**, 607–637.
McDonald, R.P. and Marsh, H.W. (1990) Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin*, **107**, 247–255.
Morgan, R.B. and Smith, J.E. (1996) *Staffing the new workplace: Selecting and promoting for quality improvement*. Illinois: ASQC Quality Press.
Morris, L., Davis, M. and Hutchings, C. (1981) Cognitive and emotional components of anxiety: Literature review and a revised Worry-Emotionality Scale. *Journal of Educational Psychology*, **73**, 541–555.
Nottelmann, E.D. and Hill, D.T. (1977) Test anxiety and off-task behavior in evaluative- situations. *Child Development*, **48**, 225–231.
Oswald, D.L. and Harvey, R.D. (2000) Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, Personality, Social*, **19**, 338–356.
Plass, J.A. and Hill, K.T. (1986) Children's achievement strategies and test performance: The role of time pressure, evaluation anxiety, and sex. *Developmental Psychology*, **22**, 31–36.
Powers, D.E. (1986) *Test anxiety and the GRE general test*. Report No. 86-45 Princeton, NJ: Educational Testing Service.
Ross, E.W. (1999) Resisting test mania. *Theory and Research in Social Education*, **27**, 126–127.
Rothstein, H.R. and McDaniel, M.A. (1992) Differential validity by sex in employment settings. *Journal of Business and Psychology*, **65**, 428–435.
Ryan, A.M. and Ployhart, R.E. (2000) Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, **26**, 565–606.
Ryan, A.M., Ployhart, R.E., Greguras, G.J. and Schmit, M.J. (1998) Test preparation programs in selection contexts: Self-selection and program effectiveness. *Personnel Psychology*, **51**, 599–621.
Rynes, S.L., Bretz, R. and Gerhart, B. (1991) The importance of recruitment in job choice: A different way of looking. *Personnel Psychology*, **44**, 487–521.
Sarason, I.G. (1959) Intellectual and personality correlates of test anxiety. *Journal of Abnormal and Social Psychology*, **66**, 73–75.
Sarason, I.G. (1984) Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology*, **46**, 929–938.
Schat, A.C. and Hausdorf, P.A. (2000, July) Differential validity of ability tests by sex in a Canadian transit organization: Some unexpected findings. Paper presented at the ASAC-IFSAM Conference, Montreal, Quebec.
Schmidt, F.L. and Hunter, J.E. (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.
Schmitt, N., Mellon, P.M. and Bylenga, C. (1978) Sex differences in validity for academic and employment criteria, and different types of predictors. *Journal of Applied Psychology*, **63**, 145–150.
Schmit, M.J. and Ryan, A.M. (1992) Test-taking dispositions: A missing link? *Journal of Applied Psychology*, **77**, 629–637.
Schmit, M.J. and Ryan, A.M. (1997) Applicant withdrawal: The role of test-taking attitudes and racial differences. *Personnel Psychology*, **50**, 855–867.
Seipp, B. (1991) Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, **4**, 27–41.
Sipos, K., Sipos, M. and Spielberger, C.D. (1986) First results with the Hungarian test anxiety inventory. In C.D. Spielberger and R. Diaz-Guerrero (Eds.), *Cross Cultural Anxiety* (4th Edn.). Washington: Hemisphere Publishing.
Smith, K.L., Michael, W.B. and Hocevar, D. (1990) Performance on creativity measures with examination-taking instructions to induce high or low levels of test anxiety. *Creativity Research Journal*, **3**, 265–280.
Spielberger, C.D. (1972) Conceptual and methodological issues in anxiety research. In C.D. Speilberger (Ed.), *Anxiety: Current Trends in Theory and Research* (Vol. 2, pp. 481–493). New York: Academic Press.
Spielberger, C.D. (1980) *Test Anxiety Inventory Manual*. California: Mind Garden.
Spielberger, C.D. and Vagg, P.R. (1995) *Test anxiety: Theory, assessment, and treatment*. Washington, DC: Taylor and Francis.
Steele, C.M. (1997) A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, **52**, 613–629.
Steiger, J.H. (1989) *Causal modeling: A supplementary module for SYSTAT and SYGRAPH*. Evanston IL: Systat.
Swarthout, D., Synk, D. and Goode, W. (1984) *The effects of sex on General Aptitude Test Battery validity and test scores*. Test Research Report No. 49. US Department of Labor. Washington, DC.
Tamres, L.K., Janicki, D. and Helgeson, V.S. (2002) Sex differences in coping behavior: A meta-analytic review and an examination of relative coping. *Personality and Social Psychology Review*, **6**, 2–30.
Tett, R.P., Jackson, D.N., Rothstein, M. and Reddon, J.R. (1994) Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick, and Hunter. *Personnel Psychology*, **47**, 157–172.
Viswesvaran, C. and Ones, D.S. (1995) Theory testing: Combining psychometric meta-analysis and structural equation modeling. *Personnel Psychology*, **48**, 865–885.
Walsh, M., Hickey, C. and Duffy, J. (1999) Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles*, **41**, 219–240.
Zeidner, M. (1991) Test anxiety and aptitude test performance in an actual college admission testing situation: Temporal considerations. *Personality and Individual Differences*, **2**, 101–109.
Zeidner, M. (1998) *Test anxiety: The state of the art*. New York: Plenum Press.