Factorial Invariance of the Statistical Anxiety Rating Scale across Sex and Students’ Classification

Eric D Teman, Ph.D., University of Wyoming

Available at: https://works.bepress.com/ericteman/1/
Factorial invariance of the statistical anxiety rating scale across sex and students’ classification

Eric D. Teman
Department of Applied Statistics and Research Methods
University of Northern Colorado

Summary

The Statistical Anxiety Rating Scale (STARS) was used to measure statistics anxiety across 423 graduate and undergraduate students from a midsized university, in the western United States. Students’ responses were analyzed using confirmatory factor analysis (CFA) to assess the validity of scores from the proposed six-factor model, which was well-fitting, according to various adjunct fit indexes. Students’ responses were then examined using multigroup CFA to explore factorial invariance across sex and student classification (i.e., undergraduates and graduates). The model was found to be factorially invariant across sex, but not across student classification, possibly meaning graduate and undergraduate students ascribed different meaning to some items. If one ignores the test of factorial invariance, between-groups statistical tests can be unduly influenced by measurement artifacts, sometimes erroneously identifying statistically significant mean differences when there are none.

Many university students taking statistics classes report statistics anxiety (Onwuegbuzie, 2004). Statistics anxiety has been researched for several decades (Earley & Mertler, 2002), using numerous tests. One of the most popular tests is the Statistical Anxiety Rating Scale (STARS; Cruise, Cash, & Bolton, 1985). The STARS was developed to measure statistics anxiety using a sample of 423 students (graduate and undergraduate) across a variety of academic disciplines. Statistics anxiety has been defined in several ways, but commonly cited ones are attributable to Cruise, et al. (1985), Zeidner (1990), and Onwuegbuzie, Da Ros, and Ryan (1997). Onwuegbuzie, et al. defined statistics anxiety as “a state-anxiety reaction to any situation in which a student is confronted with statistics in any form and at any time” (p. 28). Cruise, et al.’s definition stated “the feelings of anxiety encountered when taking a statistics course or doing statistical analysis; that is, gathering, processing, and interpreting” (p. 92), whereas Zeidner defined statistics anxiety as a performance characterized by extensive worry, intrusive thoughts, mental disorganization, tension, and physiological arousal … when exposed to statistics content, problems, instructional situations, or evaluative contexts, and is commonly claimed to debilitate performance in a wide variety of academic situations by interfering with the manipulation of statistics data and solution of statistics problems. (p. 319)

Prior estimates show statistics anxiety is experienced by the majority of graduate students at uncomfortable levels (Onwuegbuzie, 2004). Of major concern is that performance in a statistics class and magnitude of statistics anxiety are negatively related (Zeidner, 1990; Onwuegbuzie & Seaman, 1995; Fitzgerald & Jurs, 1996). This is alarming, considering the important role statistics plays in quantitative research (Birenbaum & Eylath, 1994). Graduate students should be able to readily interpret statistical findings in scholarly publications (Birenbaum & Eylath, 1994). Knowledge of statistics and applying statistical techniques are ever more critical in all academic disciplines (Baloglu, 2003; Mji, 2009). Although research methods are commonly taught separately from statistics in graduate programs, research methods per se are not typically part of the undergraduate curriculum, so statistics courses may be students’ only formal introduction to research methods. Statistics anxiety is not the principal concern, but rather the outcomes related to anxiety (Mji, 2009). Anxious students may have difficulty learning and using statistics.
There are many explanations for possible sources of statistics anxiety. These are often broken down into three categories: environmental, dispositional, and situational (Baloğlu, 2003). Environmental aspects might include sex, age, ethnicity, academic major, and previous mathematics experiences (Baloğlu, 2003), which may be described as the biases one brings into the statistics course (Onguegbuzie, et al., 1997). Some prior research indicated women experience difficulty in quantitative areas (Royse & Romph, 1992) and that women experienced higher statistics anxiety (Zeidner, 1990; Onguegbuzie & Seaman, 1995); however, other research has indicated that no such significant differences (Cruise & Wilkins, 1980; Baloğlu, 2003). Unlike environmental factors, situational factors associated with greater of statistics anxiety have been reported during enrollment in statistics class. These factors include exposure to statistical definitions, the instructor (Onguegbuzie, et al., 1997), lack of feedback from the instructor (Zeidner, 1991), and the general nature of the statistics class. Dispositional factors which might be related to statistics anxiety include learning styles (Onguegbuzie, 1998), general attitude toward statistics (Harvey, Plake, & Wise, 1985), and perceptions of statistics (Zeidner, 1991). For instance, because many students dread being required to take a statistics course, they often take these courses at the end of their degree program (Onguegbuzie & Wilson, 2003; Zeidner, 1991). Waiting until the end of their academic careers before enrolling in a required statistics class means not actively applying statistics during their academic training (Onguegbuzie, et al., 1997). In addition, many students perceive statistics as the most difficult class they have to take (Schacht & Stewart, 1990) and perform worse than in other classes (Onguegbuzie, Slate, Paterson, & Watson, 2000).

There is also evidence of higher statistics anxiety among graduate students than undergraduates (Harvey, et al., 1985; Benson & Bandalos, 1989). On the other hand, in a separate study, Benson (1989) reported statistics anxiety did not differ statistically significantly between undergraduate and graduate students. Higher statistics anxiety also was correlated negatively with the length of the class (Bell, 2001), suggesting shorter courses may be linked to higher anxiety.

Psychometrics of the STARS

Factorial invariance (also known as equivalence of measurement) allows the assumption that a construct holds the same meaning for the different groups tested. It is proposed that by ignoring a test of factorial invariance, researchers essentially ignore a fundamental measurement assumption when conducting between-group comparisons. Then, a comparison could be statistically significant or not only as a result of a measurement artifact, namely nonequivalence of measurement between groups. The implications of ignoring this issue in the context of the STARS should be examined. Even researchers not interested in statistics anxiety may have interest in the broad implications of measurement non-equivalence, which, in the extreme, could put into question one’s statistical conclusions.

Baloğlu (2002) attempted to confirm the six-factor model of STARS using a sample of 221 undergraduate college students. All factor loadings were significant (p < .05). Model fit was also assessed using several fit indices, including the goodness-of-fit index (GFI; Jöreskog & Sörbom, 1982), comparative fit index (CFI; Bentler, 1990), and root mean square error of approximation (RMSEA; Steiger, 1990). All measures of fit indicated the six-factor model was not a good fit for the data, so support for construct validity of the STARS was minimal in this group. Hanna, Shevlin, and Dempster (2008) conducted confirmatory factor analyses, testing one-, four-, and six-factor models with 849 undergraduate psychology students in the United Kingdom. Appropriate fit indices, including RMSEA, CFI, and standardized root mean square residual (SRMR) indicated the one-factor model fit the data poorly. Both the 4- and six-factor models exhibited reasonable model fit, with the six-factor model fitting the data the best.

Dauphinee, Schau, and Stevens (1997) conducted a confirmatory factor analysis of the Survey of Attitudes Toward Statistics (SATS) in a group of 991 undergraduate students. An invariance analysis across sex suggested that the SATS model was equivalent across sex. Maximum likelihood confirmatory factor analysis was used using LISREL 7. Although this was then the only known method of performing an invariance analysis, this method has now been judged to be inaccurate for use with ordinal data. Instead, a weighted least squares mean- and variance-adjusted (WLSMV) estimator must be used to accommodate ordinal data (Lubke & Muthén, 2004; Millsap & Yun-Tein, 2004). This estimator is generally robust under ordinal (non-normal) data conditions (Flora & Curran, 2004; Hutchinson, Raymond, & Black, 2008).

Clearly, statistics anxiety has been studied for decades; but, there are no studies which have first assessed whether the validity of the scores from STARS are equivalent across different subgroups. This is a particularly important issue because absence of measurement equivalence implies that group responses are not meaningfully comparable (Byrne, Shavelson, & Muthén, 1989; Vandenberg & Lance, 2000). Absence of measurement equivalence prohibits valid score comparison across different subgroups because the comparison is essentially “misleading and illegitimate” (Hui & Triandis, 1985, p. 134). Without measurement equivalence, notable mean differences in statistics anxiety across different groups may be attributable to measurement
Factorial Invariance / E. D. Teman

artifacts rather than real differences in perception of statistics anxiety (Hutchinson, et al., 2008). If there is no measurement equivalence across comparison groups, it is possible that prior research results are invalid because the assumption of equivalent groups was incorrect. The current study brings to the forefront the importance of measurement equivalence when making group comparisons. For instance, if subpopulations interpret the meaning of the STARS items differently, then no accurate comparison of groups can be meaningfully made.

The research on statistics anxiety is fairly extensive, but there are clear weaknesses in the literature. Although the STARS has been a popular scale for measuring statistics anxiety, there has been no psychometric research on whether STARS is measuring statistics anxiety equivalently for all students. Perhaps previous studies have been published under the assumption that STARS does measure statistics anxiety equivalently for all students. This assumption may be incorrect. Therefore, the rationale for the present study was to assess whether measurement is equivalent among different groups of students who have typically been compared.

Hypothesis 1. Statistics anxiety will exhibit factorial invariance across sex.

Hypothesis 2. Statistics anxiety will exhibit factorial invariance across students’ year of study.

Method

Participants

Upon IRB approval, students at a midsized university in the Western part of the USA were recruited from intact classrooms to complete the STARS and a demographic questionnaire. There were 423 participants (293 women, Mage = 26.6 yr., SD = 10.5; 130 men, Mage = 24.9 yr., SD = 9.6). Both undergraduate students (120 freshmen, 64 sophomores, 51 juniors, 23 seniors) and graduate students (53 master’s, 112 doctoral) were recruited to take the survey. Students were enrolled in introductory undergraduate statistics and psychology classes, or in intermediate and advanced graduate statistics and research methods classes. Students across various academic disciplines were represented in this sample, including behavioral sciences (n = 89), business (n = 14), education (n = 41), health sciences (n = 82), performing and visual arts (n = 6), social sciences (n = 31), and other (n = 8).

Measures

The survey included the STARS instrument by demographic questions (see Appendix, pp. ***). Each of the 51 items in the STARS (Cruise & Wilkins, 1980) was rated on a 5-point scale. For the first 23 items measuring test and class anxiety, interpretation anxiety, and fear of asking for help, participants rated their anxiety using anchors of 1: No anxiety and 5: Strong anxiety. A sample item is “Studying for an examination in a statistics course”. For the next 28 items measuring worth of statistics, computational self-concept, and fear of statistics teachers, participants were asked to rate their agreement using anchors of 1: Strongly agree and 5: Strongly disagree. A sample item is “Statistics is worthless to me since it is empirical and my area of specialization is abstract”. For Items 1 to 23, higher scores on each item correspond to higher anxiety; and for Items 24 to 51, higher scores on each item correspond to more positive attitudes.

Onwuegbuzie (1999) reported an estimate of internal reliability of scores on the STARS of .78 on the Worth of Statistics subscale and .84 on the Test and Class Anxiety subscale with a median of .8 for 225 African-American participants at a small suburban college of education in a mid-southern state. Baloğlu (2002) reported internal consistency reliability coefficients on the STARS scores of .64 on the Fear of Statistics Teachers subscale of .94 on the Worth of Statistics subscale for 246 college students.

The survey also comprised a series of demographic questions, including sex, age, academic major, classification (i.e., freshman, sophomore, junior, senior, master’s, doctoral), mathematics background (i.e., statistics course in high school, undergraduate statistics course in college, graduate statistics courses, high school algebra, college algebra, trigonometry, and calculus). For mathematics background, students could check all that applied.

Procedure

Participants in undergraduate introductory statistics courses were given a hard copy of the survey and a separate sheet of paper with the URL to access the online version. The class was given the option of either taking the survey in paper format or online; all chose to complete the hard copy. Participants in another class were provided a hard copy of the survey without an option to complete online. Other classes were only provided with the URL to access the online survey. The online survey format was also used as a follow-up method for collecting data—an e-mail was sent out into introductory statistics and psychology students with the URL of the survey. In all, 69 participants completed the paper survey and 354 completed the online survey.

The human participants consent form was on the first page of the paper survey packet. Participants were explicitly told to tear off the consent form to keep for their own records. The next six pages comprised the STARS. A cash prize entry form was attached as the last page; participants were told to detach this page from their own records. The next six pages comprised the STARS. A cash prize entry form was attached as the last page; participants were told to detach this page from the survey packet and fill in their name, phone number, and e-mail address if they were interested in having a chance to win the $20 cash prize. The cash prize entry forms were kept in a separate stack from the rest of the survey. Entry forms were shuffled to minimize possible
matching of entry form to participants’ corresponding surveys.

Those participants who chose to complete the survey online were first presented a screen encompassing the human participant consent form. At the bottom of the screen were two options: I consent and I decline. Declining to consent directed the participant to the university’s website. Consenting directed the participant to the survey. Online, if a participant wished to enter into the $20 cash prize drawing, a text box was available to provide the necessary contact information (i.e., name, phone number, e-mail address). If the participant felt his anonymity could be compromised by entering contact information into the textbox, an e-mail address was also provided for separate transmission.

Before data collection began, an application for exempt review was submitted to and subsequently approved by the institutional review board at the university at which the study was conducted. Depending on the feasibility of the researcher personally being able to distribute paper surveys, some students were only given the option to take the online version. For all introductory statistics students, a follow-up email was sent to indicate that they could take the survey online if they had not taken it in class. The primary purpose of this follow-up was to solicit a response from those students who did not attend class on the day the paper survey was distributed.

RESULTS

Preliminary Analyses

Data were entered into PASW 18 (IBM, 2010) and preliminary descriptive (e.g., means, standard deviations, frequencies) and internal consistency reliability analyses (e.g., Cronbach’s alpha) were run. Cronbach’s alpha was .96 for the total score. Reliability coefficients for the six subscales are shown in Table 1, which are relatively consistent with estimates found in prior studies.

Baseline Model

A six-factor confirmatory factor analysis model was specified with all observed variables being ordered categorical and estimated with Mplus (Version 5.21; Muthén & Muthén, 2010), using the WLSMV estimator. Overall model fit and component fit were examined using several measures of fit, including an appropriate $\chi^2$ test employing the WLSMV estimator, RMSEA, Tucker-Lewis index (TLI; Tucker & Lewis, 1973), and CFI.

The six-factor model in the current study fit well. Although the $\chi^2$ test statistic was statistically significant, as shown in Table 2, several adjunct fit indexes provided evidence supporting the six-factor model. The CFI was .943 and the TLI was .940 (each with a range of 0 to 1.0), where the closer the measure is to 1.0 the better the model fit. The RMSEA was .059, where a lower number is more desirable and indicates better fit (RMSEA values can range from 0 to infinity and the RMSEA is considered a measure of “badness of fit”). These fit values were compared to Hu and Bentler’s (1999) guidelines for cutoff criteria and indicate adequate model fit. The RMSEA was higher than the suggested cutoff value of .05 for close fit (Browne & Cudeck, 1993); however, Brown and Cudeck also stated that a RMSEA value less than .08 is indicative of reasonable fit. The CFI was slightly lower than .95, which is the cutoff criterion suggested by Hu and Bentler (1999) and indicates model fit may not be ideal. In regard to component fit, all parameter estimates (standardized factor loadings) were statistically significant ($p < .01$) and in the expected direction, which is indicative of good fit.

Model modification indices (MI) were examined to identify areas of misfit, which provides an estimate of

<table>
<thead>
<tr>
<th>Scale</th>
<th>Women (n = 293)</th>
<th>Men (n = 130)</th>
<th>Undergraduate (n = 258)</th>
<th>Graduate (n = 165)</th>
<th>Total (N = 423)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test and Class Anxiety</td>
<td>.87</td>
<td>.88</td>
<td>.87</td>
<td>.90</td>
<td>.88</td>
</tr>
<tr>
<td>Interpretation Anxiety</td>
<td>.88</td>
<td>.87</td>
<td>.89</td>
<td>.88</td>
<td>.88</td>
</tr>
<tr>
<td>Fear of Asking for Help</td>
<td>.84</td>
<td>.83</td>
<td>.82</td>
<td>.87</td>
<td>.84</td>
</tr>
<tr>
<td>Worth of Statistics</td>
<td>.95</td>
<td>.95</td>
<td>.94</td>
<td>.92</td>
<td>.95</td>
</tr>
<tr>
<td>Fear of Statistics Teacher</td>
<td>.80</td>
<td>.76</td>
<td>.80</td>
<td>.77</td>
<td>.79</td>
</tr>
<tr>
<td>Computational Self-concept</td>
<td>.87</td>
<td>.86</td>
<td>.86</td>
<td>.87</td>
<td>.86</td>
</tr>
<tr>
<td>Total Scale Score</td>
<td>.96</td>
<td>.96</td>
<td>.96</td>
<td>.96</td>
<td>.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>$\chi^2$</td>
<td>df</td>
<td>$p$</td>
<td>TLI</td>
<td>CFI</td>
</tr>
<tr>
<td>Initial model (51 items)</td>
<td>2965.58</td>
<td>1209</td>
<td>&lt;.001</td>
<td>.940</td>
<td>.943</td>
</tr>
<tr>
<td>Revised model (49 items)</td>
<td>2561.61</td>
<td>1112</td>
<td>&lt;.001</td>
<td>.947</td>
<td>.950</td>
</tr>
</tbody>
</table>
the decrease in $\chi^2$ for the overall model if a given parameter were freed for estimation (Brown, 2006). Two items from the Agreement subscale produced extreme modification index values on the factor loadings. Item 11, “Since I have never enjoyed math I do not see how I can enjoy statistics,” was set to load on the factor computational self-concept, and Item 24, “Statistical figures are not fit for human consumption,” were set to load on the factor Worth of Statistics. Given possible multiple factor loadings suggested by the modification indices for these two items, they were removed from the model and the CFA was re-run. The fit of the new model was slightly better, as shown in Table 2. Table 2 summarizes the fit statistics for the two versions of the baseline model. A baseline model is the first step for subsequent model comparisons.

**Invariance Results**

**Invariance by sex.**—Before testing for invariance across sex, the six-factor baseline model was separately fit for women and men to assess whether the model provided good fit for each group separately. Adjunct fit indexes suggested good global model fit for both sexes (Table 3). In regard to component fit, all parameter estimates were statistically significant for both sexes and indicative of good fit.

The first step of a factorial invariance analysis is to assess configural invariance, which requires identical models to be specified for both men and women with the added stipulation that parameters are separately estimated for the men and women. As shown in Table 3, despite the $\chi^2$ test being statistically significant, the CFI, TLI, and RMSEA are suggestive of good fit. Therefore, the six-factor model seems to be suitable for both men and women.

The second step of the invariance analysis, metric invariance (Vandenberg & Lance, 2000), tested what the decrement in model fit would be statistically significant if the factor loadings were constrained equal for the men and women. The $\chi^2$ difference test was not statistically significant $p = .02$ (appling Bonferroni adjustment and using alpha of .01 to account for numerous statistical tests), indicating items appear to be functioning similarly for both groups. Therefore, the invariance testing continued.

The test for scalar invariance (Vandenberg & Lance, 2000) was the third step in the invariance analyses, where it was tested whether the item thresholds were invariant across men and women. The $\chi^2$ difference test was not statistically significant (Table 3), meaning scalar invariance held, so the invariance testing proceeded to the test of latent means differences.

Latent mean differences between men and women for both the Test and Class Anxiety factor and the Interpretation Anxiety factor were statistically significant (Table 4), indicating that women were reporting higher anxiety in those areas. Differences between latent means were not statistically significant for the men and women on Fear of Asking for Help, Worth of Statistics, Fear of Statistics Teacher, or Computational Self-concept (Table 4).

**Invariance by students’ classification.**—Separate CFA models were tested for undergraduate and graduate students. According to the adjunct fit indexes (Table 3), the baseline model fit adequately for undergraduates; however, both CFI and TLI were slightly lower than

### TABLE 3

<table>
<thead>
<tr>
<th>Education Level</th>
<th>Model</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$\Delta\chi^2$</th>
<th>$\Delta df$</th>
<th>$p$</th>
<th>TLI</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Men</td>
<td>1478.04</td>
<td>1112</td>
<td></td>
<td>&lt;.0001</td>
<td></td>
<td>.955</td>
<td>.957</td>
<td>.050</td>
<td></td>
</tr>
<tr>
<td>Baseline Women</td>
<td>2181.14</td>
<td>1112</td>
<td></td>
<td>&lt;.011</td>
<td></td>
<td>.944</td>
<td>.947</td>
<td>.057</td>
<td></td>
</tr>
<tr>
<td>Configural</td>
<td>3124.37</td>
<td>1956</td>
<td></td>
<td>&lt;.001</td>
<td></td>
<td>.952</td>
<td>.955</td>
<td>.053</td>
<td></td>
</tr>
<tr>
<td>Metric</td>
<td>3164.07</td>
<td>1996</td>
<td>60.11</td>
<td>40</td>
<td>.02</td>
<td>.953</td>
<td>.955</td>
<td>.053</td>
<td></td>
</tr>
<tr>
<td>Scalar</td>
<td>3245.82</td>
<td>2119</td>
<td>153.95</td>
<td>123</td>
<td>.03</td>
<td>.957</td>
<td>.956</td>
<td>.050</td>
<td></td>
</tr>
<tr>
<td>Baseline Undergraduate</td>
<td>1978.40</td>
<td>1112</td>
<td></td>
<td>&lt;.011</td>
<td></td>
<td>.939</td>
<td>.942</td>
<td>.055</td>
<td></td>
</tr>
<tr>
<td>Baseline Graduate</td>
<td>1533.96</td>
<td>1112</td>
<td></td>
<td>&lt;.001</td>
<td></td>
<td>.952</td>
<td>.955</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td>Configural</td>
<td>3166.65</td>
<td>2044</td>
<td></td>
<td>&lt;.001</td>
<td></td>
<td>.950</td>
<td>.952</td>
<td>.051</td>
<td></td>
</tr>
<tr>
<td>Metric</td>
<td>3248.82</td>
<td>2085</td>
<td>142.51</td>
<td>41</td>
<td>&lt;.001</td>
<td>.949</td>
<td>.951</td>
<td>.051</td>
<td></td>
</tr>
</tbody>
</table>

*Based on WLSMV-adjusted calculations.

## TABLE 4

<table>
<thead>
<tr>
<th>Factor</th>
<th>Latent Mean</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test and Class Anxiety</td>
<td>-0.405</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interpretation Anxiety</td>
<td>-0.422</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fear of Asking for Help</td>
<td>-0.252</td>
<td>.04</td>
</tr>
<tr>
<td>Worth of Statistics</td>
<td>-0.021</td>
<td>.85</td>
</tr>
<tr>
<td>Fear of Statistics Teacher</td>
<td>0.170</td>
<td>.19</td>
</tr>
<tr>
<td>Computational Self-concept</td>
<td>-0.062</td>
<td>.60</td>
</tr>
</tbody>
</table>

*Note.*—A negative latent mean indicates that women exhibited higher anxiety or more agreement.
desirable. Estimates of all parameters were statistically significant. The CFA model for graduate students fit well (Table 3). Estimates of all parameters were statistically significant for both groups.

Configural invariance was observed, despite a statistically significant $\chi^2$ value (Table 3), indicating support for the tenability of the same pattern of fixed and free factor loadings across undergraduate and graduate students. Metric invariance, however, did not hold across this subpopulation, with the $\chi^2$ difference test being statistically significant (Table 3). In general, lack of metric invariance indicates that some items may have behaved differently across undergraduate students and graduate students, even though the overall factor structure held across the two groups.

Because factorial invariance failed at the metric level of testing, scalar invariance was not tested. Instead, post hoc tests were performed to study the nature of between-group differences for the latent variable. Chi-square difference tests were conducted for the six latent variables (Table 5). From the statistical tests, one may infer both configural and metric invariance were evident for two factors, Test and Class Anxiety and Interpretation Anxiety. The factor Fear of Asking for Help passed the test for configural, metric, and scalar invariance. However, the last three factors (Worth of Statistics, Fear of Statistics Teachers, Computational Self-concept) were noninvariant on the factor loadings (metric invariance), which may suggest of undergraduate and graduate students interpreting the STARS items in different ways.

### DISCUSSION

Cruise and other authors using the STARS suggested that the construct of statistics anxiety consists of six factors: Worth of Statistics, Interpretation Anxiety, Test and Class Anxiety, Computational Self-concept, Fear of Asking for Help, and Fear of Statistics Teachers. A CFA confirmed the tenability of this six-factor structure. The WLSMV estimator was used because of its robustness to handle ordinal data (Flora & Curran, 2004).

Use of the WLSMV estimator does not seem to be described in studies examining rating scale data from scales purporting to measure statistics anxiety. Results of studies not employing the WLSMV estimator were based on state-of-the-art estimators at the time; however, new evidence in the form of a more appropriate estimator could bring the prior results in question. For instance, Dauphinee, et al. (1997) analyzed rating scale data from participants complete the Survey of Attitudes Towards Statistics (SATS) employing the maximum likelihood estimator to conduct a multigroup CFA across sex and found the measurement model to be equivalent across both groups. However, it is conceivable that by employing the WLSMV estimator, which would have produced unconfounded estimates of thresholds and factor loadings (Lubke & Muthén, 2004; Millsap & Yun-Tein, 2004), may have led to different conclusions.

In the current study, the factorial invariance tests in the current study employed the WLSMV estimator in the multigroup CFA and indicated that the six-factor measurement model was equivalent across sex. The

### TABLE 5

<table>
<thead>
<tr>
<th>Posthoc Model Fit for Student Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td><strong>Test and Class Anxiety</strong></td>
</tr>
<tr>
<td>Configural</td>
</tr>
<tr>
<td>Metric</td>
</tr>
<tr>
<td>Scalar</td>
</tr>
<tr>
<td><strong>Interpretation Anxiety</strong></td>
</tr>
<tr>
<td>Configural</td>
</tr>
<tr>
<td>Metric</td>
</tr>
<tr>
<td>Scalar</td>
</tr>
<tr>
<td><strong>Fear of Asking for Help</strong></td>
</tr>
<tr>
<td>Configural</td>
</tr>
<tr>
<td>Metric</td>
</tr>
<tr>
<td>Scalar</td>
</tr>
<tr>
<td><strong>Worth of Statistics</strong></td>
</tr>
<tr>
<td>Configural</td>
</tr>
<tr>
<td>Metric</td>
</tr>
<tr>
<td><strong>Fear of Statistics Teacher</strong></td>
</tr>
<tr>
<td>Configural</td>
</tr>
<tr>
<td>Metric</td>
</tr>
<tr>
<td><strong>Computational Self-concept</strong></td>
</tr>
<tr>
<td>Configural</td>
</tr>
<tr>
<td>Metric</td>
</tr>
</tbody>
</table>
presence of metric invariance indicates that the individual items from the STARS function similarly for men and women. Scalar invariance indicates the thresholds do not differ by sex.

Because the model by sex was invariant across the thresholds, tests of latent means were performed for each of the six factors. Two sex differences were identified as statistically significant: Women had higher mean ratings on Test and Class Anxiety and Interpretation Anxiety. Further implications of invariance of the thresholds include the validity of between-group comparisons made for sex. If the researcher compares observed (or latent means) on two groups, an independent-samples $t$ test would be meaningful and readily interpretable as a true mean difference. However, in the absence of metric invariance, researchers should be cautious in interpreting the meaning of such differences.

On the other hand, invariance tests across undergraduate and graduate students indicated that these two groups may not have ascribed the same meaning to all items on the STARS, i.e., scalar noninvariance was present. This finding precluded any meaningful between-group comparisons of these two groups (Vandenbogen & Lance, 2000). In other words, the STARS should not be used if the purpose is to compare mean scores of undergraduate and graduate students.

Although some research supports higher scores on statistics anxiety among graduate students than for undergraduate students (e.g., Benson & Bandalos, 1989; Harvey, et al., 1985), tests for factorial invariance were not conducted. Therefore, it is unknown whether the results are meaningful for the samples in those studies.

Conclusions
Although discussions of equivalence of measurement or factorial invariance may not be found in many statistics texts, researchers should be aware of the effect of imprecise measurement in statistical analyses. In particular, lack of factorial invariance of metrics (thresholds) should preclude between-groups comparisons, even for an independent-samples $t$ test. For instance, should one find a statistically significant difference in observed (or latent) sample means without first conducting a multigroup CFA to test for the presence of factorial invariance, there can be no certainty that this difference reflects a real difference in the trait of interest. It is possible the difference is an artifact of measurement, reflecting nonequivalence in the sample.

Factorial invariance should be viewed as an assumption in between-groups statistical tests, as one should strengthen the validity of statistical conclusions based on comparison of group scores on a trait. Before assuming that certain subpopulations of interest are invariant, that assumption must first be tested to ensure such comparisons are valid.

Limitations
The sample of students in the current study came from one university and therefore may affect generalizability of results to other populations. Invariance tests by academic discipline were not conducted in the current study, as group sample sizes were not sufficient. Future research should focus on obtaining larger and more diverse samples by academic concentration, as statistics anxiety might manifest differently for physical science majors as opposed to liberal arts majors, for example.

References
Factorial Invariance / E. D. Teman

APPENDIX

Statistics Attitude Survey: Part 1 (23 items)
On the first 23 items, indicate how much anxiety you experience in each of the situations on a scale from 1 to 5 (where 1=no anxiety and 5=strong anxiety).

1. Studying for an examination in a statistics course
   No anxiety 1 2 3 4 5 Strong anxiety

2. Interpreting the meaning of a table in a journal article
   No anxiety 1 2 3 4 5 Strong anxiety

3. Going to ask my statistics teacher for individual help with material I am having difficulty understanding
   No anxiety 1 2 3 4 5 Strong anxiety

4. Doing the homework for a statistics course
   No anxiety 1 2 3 4 5 Strong anxiety

5. Making an objective decision based on empirical data
   No anxiety 1 2 3 4 5 Strong anxiety

6. Reading a journal article that includes some statistical analyses
   No anxiety 1 2 3 4 5 Strong anxiety

7. Trying to decide which analysis is appropriate for my research project
   No anxiety 1 2 3 4 5 Strong anxiety

8. Doing an examination in a statistics course
   No anxiety 1 2 3 4 5 Strong anxiety

9. Reading an advertisement for an automobile which includes figures on gas mileage, compliance with population regulations, etc.
   No anxiety 1 2 3 4 5 Strong anxiety

10. Walking into the classroom to take a statistics test
    No anxiety 1 2 3 4 5 Strong anxiety

11. Interpreting the meaning of a probability value once I have found it
    No anxiety 1 2 3 4 5 Strong anxiety

12. Arranging to have a body of data put into the computer
    No anxiety 1 2 3 4 5 Strong anxiety

13. Finding that another student in class got a different answer than I did to a statistical problem
    No anxiety 1 2 3 4 5 Strong anxiety

14. Determining whether to reject or retain the null hypothesis
    No anxiety 1 2 3 4 5 Strong anxiety

15. Waking up in the morning on the day of a statistics test
    No anxiety 1 2 3 4 5 Strong anxiety

16. Asking one of your professors for help in understanding a printout
    No anxiety 1 2 3 4 5 Strong anxiety

17. Trying to understand the odds in a lottery
    No anxiety 1 2 3 4 5 Strong anxiety

18. Watching a student search through a load of computer printouts from his/her research
    No anxiety 1 2 3 4 5 Strong anxiety

19. Asking someone in the computer lab for help in understanding a printout
    No anxiety 1 2 3 4 5 Strong anxiety

20. Trying to understand the statistical analyses described in the abstract of a journal article
    No anxiety 1 2 3 4 5 Strong anxiety

21. Enrolling in a statistics course
    No anxiety 1 2 3 4 5 Strong anxiety

22. Going over a final examination in statistics after it has been marked
    No anxiety 1 2 3 4 5 Strong anxiety

23. Asking a fellow student for help in understanding a printout
    No anxiety 1 2 3 4 5 Strong anxiety

Statistics Attitude Survey: Part 2 (28 items)
On the next 28 items, indicate your level of agreement with each statement on a scale from 1 to 5 (where 1=strongly agree anxiety and 5=strongly disagree).

1. I am a subjective person, so the objectivity of statistics is inappropriate for me
   Strongly agree 1 2 3 4 5 Strongly disagree

2. I have not had math for a long time. I know I will have problems getting through statistics
   Strongly agree 1 2 3 4 5 Strongly disagree
3. I wonder why I have to do all these things in statistics when in actual life I will never use them  
   | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

4. Statistics is worthless to me since it is empirical and my area of specialization is abstract  
   | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

5. Statistics takes more time than it is worth  
   | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

6. I feel statistics is a waste  
   | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

7. Statistics teachers are so abstract they seem inhuman  
   | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

8. I cannot even understand seventh- and eighth-grade math; how can I possibly do statistics?  
   | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

9. Most statistics teachers are not human  
   | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

10. I lived this long without knowing statistics, why should I learn it now?  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

11. Since I have never enjoyed math I do not see how I can enjoy statistics  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

12. I do not want to learn to like statistics  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

13. Statistics is for people who have a natural leaning toward math  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

14. Statistics is a pain I could do without  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

15. I do not have enough brains to get through statistics  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

16. I could enjoy statistics if it were not so mathematical  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

17. I wish the statistics requirement would be removed from my academic program  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

18. I do not understand why someone in my field needs statistics  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

19. I do not see why I have to fill my head with statistics. It will have no use in my career  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

20. Statistics teachers speak a different language  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

21. Statisticians are more number oriented than they are people oriented  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

22. I cannot tell you why, but I just do not like statistics  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

23. Statistics teachers talk so fast you cannot logically follow them  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

24. Statistical figures are not fit for human consumption  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

25. Statistics is not really bad. It is just too mathematical  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

26. Affective skills are so important in my (future) profession that I do not want to clutter my thinking with something as cognitive as statistics  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

27. I am never going to use statistics so why should I have to take it?  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |

28. I am too slow in my thinking to get through statistics  
    | Strongly agree | 1 | 2 | 3 | 4 | 5 | Strongly disagree |
Statistical Attitude Survey: Part 3

1. Gender:  ____ Male  ____ Female

2. Age: ______

3. Major (choose only one):
   ___ Behavioral Sciences (e.g., counseling, psychology)
   ___ Business
   ___ Education (e.g., early childhood, elementary, secondary, HESAL)
   ___ Health Sciences (e.g., audiology, speech-language sciences, dietetics, gerontology, public health, recreation & tourism, rehabilitation, human services, sport & exercise science).
   ___ Humanities (e.g., history, philosophy, political science, English, communication, journalism)
   ___ Mathematics (e.g., mathematics, mathematics education, statistics, computer science)
   ___ Natural Sciences (e.g., biology, chemistry, physics, earth sciences)
   ___ Nursing
   ___ Performing & Visual Arts
   ___ Social Sciences (e.g., anthropology, economics, geography, sociology, criminal justice)
   ___ Other (please specify) ___________

4. Student Classification
   ___ Freshman
   ___ Sophomore
   ___ Junior
   ___ Senior

   Graduate student (indicate masters or doctorate)
      ___ Masters student
      ___ Doctoral student

5. Mathematics Background (check ALL that apply)
   ___ Statistics course(s) in high school
   ___ Undergraduate statistics course(s) in college
   ___ Graduate-level statistic course(s)
   ___ High school algebra
   ___ College algebra
   ___ Trigonometry
   ___ Calculus