Reliability and Factor Structure of the Westside Test Anxiety Scale among University Students.

Authors:
Prashanth Talwar, Associate Professor, Faculty of Cognitive Sciences & Human Development, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak Malaysia,
Shevonne Matheiken, Specialty Registrar, Dual General Adult and Old Age Psychiatry, East London NHS Foundation Trust, Julia Lee Ai Cheng, Deputy Dean, Faculty of Cognitive Sciences & Human Development, Universiti Malaysia Sarawak,
Surena Sabih, Dean, Faculty of Cognitive Sciences & Human Development, Universiti Malaysia Sarawak.

Address for Correspondence
Dr. Prashanth Talwar,
Associate Professor
Faculty of Cognitive Sciences & Human Development
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia.
E-mail: talpra61@hotmail.com.

Citation

Submitted: Aug 5, 2019; Accepted: Oct 14, 2019; Published: Oct 30, 2019

Abstract: The Westside Test Anxiety Scale (WTAS) is a widely-used screening instrument for assessing test anxiety impairments in educational settings. However, psychometric properties of this measure have not been established adequately. Therefore, the main objective of this study was to evaluate the factor structure and measurement invariance of the WTAS. Data for the cross-sectional study was collected from a convenience sample of 218 undergraduate students enrolled in a public university in Malaysia. Construct validity was investigated by Exploratory Factor Analysis (EFA). Confirmatory Factor Analysis (CFA) was used to test two models (original structure and EFA structure). Also, multi-group invariance was analyzed in order to determine the extent to which the factor structure was comparable across gender. Results of CFA indicated that both models met all the pre-established fit criteria. There was however, more support for the re-organized two-factor EFA model, with high reliability values. Multi-group CFA demonstrated measurement invariance across gender. Findings from the present study indicate that the WTAS could serve as a screening tool to detect test anxiety among students.

Key Words: Test anxiety, Factor structure, Gender invariance

Introduction:
Examinations and tests, without doubt, are an unavoidable part of every student’s educational progress. Apparently, progress in school and entrance into or promotion within a vocational field may be, in part, a function of the capacity to demonstrate abilities and aptitudes on a test or series of tests [4]. In view of that, responses to high levels of test anxiety could have important implications in the educational setting[5], especially for university students who are aiming for a degree and aspiring to join the workforce.

Test anxiety is an undesirable reaction toward evaluation [3]. Test anxiety is defined as the physiological and behavioral responses related to taking oral or written tests that is experienced before and during test taking about possible negative consequences or failure on an examination[6]. Research on test anxiety has repeatedly attempted to provide a more refined measurement of multiple dimensions of the construct [7]. On the whole, there has been general consensus that emotionality and worry constitute the two primary dimensions of test anxiety [8,9], leading to stress. Stress levels may escalate to significant proportions in some students, to present with symptoms of anxiety especially during tests and examination periods[10-13]. Academic stress can also result in increased pervasiveness of psychological and physical problems like depression, behavioral problems, irritability and procrastination, which in turn can affect their academic results.[14-18] It was demonstrated in a number of studies that parental pressures and teachers’ expectations were associated with stress around the time of examinations or about choosing particular academic study or a future career [11]. There has been considerable interest in the level of manifestation of test-anxiety as it relates to such topics as verbal or motor learning, stimulus generalization, form discrimination and size estimation [19]. In developing a measurement for assessing the level of test anxiety, Driscoll (2004) devised the Westside Test Anxiety Scale (WTAS)[20] as a concise instrument which could be completed by respondents in approximately 5-8 minutes[20].
was validated by comparing fifth graders and college-aged students, and found to have a Pearson $r = 0.44$[21]. The scale has also demonstrated the internal consistency in a study comparing different types of education styles [22]. However, despite being used widely by researchers to measure test anxiety among student populations[23-25] psychometric testing of WTAS has remained limited. Therefore, the main objective of the present study was to assess the reliability and factor structure of the WTAS among undergraduate university students.

**Methodology**

**Participants and Procedure**

The cross-sectional study was conducted among a sample comprising 218 undergraduate students of an University in Malaysia. A convenience sampling technique was implemented whilst selecting participants who were enrolled full-time at the Faculty of Cognitive Science and Human Development. Prior to assessment, students were briefed about the purpose of the study and assured that their responses would be confidential. The self-administered questionnaire was distributed during the last 10 minutes of a 2-hour class lecture. Participation was voluntary and applicable only to students who were present on the day of assessment. Signed consent was also obtained from participating students. This study received University approval.

**Materials**

The self-report questionnaire consisted of two parts. The first section included the socio-demographic characteristics of participants, such as age, gender and ethnicity. The second section comprised of the Westside Test Anxiety Scale (WTAS). Students were required to respond to the 10 WTAS items on a 5-point Likert scale extending from 1 (never or not at all true) to 5 (extremely or always true). The total score was subsequently divided by 10, in order to interpret the level of test anxiety. According to WTAS recommendations, participants who scored records less than 1.9 depicted low test anxiety, scores of 2.0 to 2.5 indicated normal or average test anxiety, 2.5 to 2.9 showed high normal test anxiety, 3.0 to 3.4 was portrayed as moderately high, 3.5 to 3.9 as high test anxiety, and 4.0 to 5.0 as extremely high anxiety.

**Statistical analysis**

Statistical Program Social Sciences (SPSS) version 23 was used to analyze and code the data in this study. Level of significance was set at 0.05, unless stated otherwise. All categorical data have been represented with numbers and percentage. Prior to conducting primary analyses, the data was examined for univariate outliers and all were found to be within range values. Data was normally distributed, hence no variable transformation was deemed necessary. Exploratory Factor Analysis (EFA) was conducted by using the extraction method of principal component analysis with varimax rotation of axes, in order to examine the structure of WTAS scale. The number of factors to be extracted was determined through a scree plot test, in combination with the conventional cut-off criteria of eigenvalues greater than one. Factorability of the data was evaluated using Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett’s Test of Sphericity (BTS). A KMO value greater than 0.80 is considered meritorious, and if the BTS value is large with a significant p value ($p<0.05$), then the data is probably factorable [26]. The criterion for factor loading was set at $\geq 0.50$, so as to suppress absolute value less than 0.50. The WTAS factor structure that emerged from EFA was further verified by Confirmatory Factor Analysis (CFA), using structural equation modelling in AMOS v. 23. Excel was utilized to calculate the nested model comparison. By means of CFA, two competing measurement models were evaluated. The first model was a two-factor structure, which was based on the original layout of items. The second model was the constructed factor structure, derived from EFA. In each analysis, the maximum likelihood estimation method was used, and covariance matrices were assessed. For each model tested, adequacy of fit was examined by considering several indices [27], which included the chi-square statistic ($\chi^2$); the Goodness-of-Fit Index (GFI); the Adjusted Goodness-of-Fit Index (AGFI); the Comparative Fit Index (CFI); the Incremental Fit Index (IFI); the Root Mean Square Error of Approximation (RMSEA); and the Akaike Information Criterion (AIC).

The chi-square statistic ($\chi^2$) was computed for both models, wherein, ideally the values should not be significant. Generally, acceptable models do not have definite cut-off points. Nonetheless, the following criteria were used to detect the goodness-of-fit for a specific model. The GFI is an absolute index of fit, wherein a value of 1 indicates perfect fit, GFI greater than 0.90 may indicate good fit, and values close to 0 points towards very poor fit. The AGFI differs from GFI, only in the fact that it adjusts for the number of degrees of freedom in the specified model, thereby addressing the issue of parsimony. AGFI values range from 0 to 1, with values close to 1 being indicative of a good fit. The CFI is an incremental fit index and is normed so that values range between 0 and 1, where higher values indicate a better fit. In general, values greater than 0.95 and 0.97 are associated with acceptable and good fit, respectively. The IFI, which is insensitive to sample size, regards values that exceed 0.90 as acceptable. The RMSEA is an absolute fit index, in which values less than 0.05 could be considered as a good fit, 0.05 to 0.08 as adequate fit, 0.08 to 0.10 as mediocre fit, and those greater than 0.10 could suggest poor fit. The AIC addresses the issue of parsimony in the assessment of model fit, with smaller values representing a better fit. Modification indices were not utilized for both models.

To assess factor invariance, multi-group CFA was performed in AMOS v. 23, using maximum likelihood estimations. Model fit of the best-factor solution was evaluated across gender, using the same aforementioned criteria of approximate fit indices. The degree of invariance was assessed by the likelihood ratio test, also known as chi-square difference test [28]. Lastly, internal consistency reliability of the scale was assessed by computing the Composite Reliability (CR), wherein values of CR $\geq 0.70$ were considered adequate [28].

**Results**

The sample comprised of 218 undergraduate students, of which 18% were males and 82% were females. Participating students were between the ages of 20 and 24 years ($M=20$; $SD=0.86$). Majority of the students were Malays (50%), 20% were Chinese, 4% were Indians, and 26% were from other races.

**Exploratory Factor Analysis**

The internal structure of WTAS originally suggested by Driscoll (2007), consisted of two factors [29]. Factor I, labelled as ‘Incapacity’ represented memory loss and poor cognitive processing, and included items 1, 4, 5, 6, 8 and 10. Factor II, which was termed ‘Worry’ signified catastrophizing, and encompassed items 2, 3, 7 and 9. Thus, a two-factor solution was examined by EFA. The KMO Test of Sampling Adequacy which was initially conducted was found to be 0.85, and the Bartlett’s Test of Sphericity was 600.21 (df =45; $p<0.05$). Also, correlation errors were not observed among the variables. Altogether, these analyses indicated the suitability of data for performing factor analysis.

EFA results indicated that two factors had eigenvalues greater than one, and these accounted for 50.66% of item variance. The first factor (Cognitive dysfunction) which comprised of items 1, 3, 4, 7 and 8 explained 40.27% variance with an eigenvalue of 4.02. The second factor (Agony) which included items 2, 5, 6, 9 and 10 explained 10.39% variance with an
eigenvalue of 1.09. Despite the fact that a two-factor solution was replicated in the EFA, differences were observed in the original structure and the EFA structure. In both models, the only common items identified in Factor I were 1, 4 and 8, while in Factor II the common items were 2 and 9. Moreover, the EFA structure had equal number of items in both factors, when compared to the original structure.

**Confirmatory Factor Analysis**

In order to identify the best factor structure of WTAS in the present sample of students, two different CFAs were conducted to test the two respective models. The goodness-of-fit statistics obtained for both models have been summarized in Table 1. In reviewing the indices, it is evident that Model 2 (EFA structure) exhibited a better fit than Model 1 (original structure), according to the standard evaluative criteria used. It has been indicated that models with smaller AIC may have better fit than models with higher AIC, which was likewise identified here. Also, bigger values for the GFI, AGFI, CFI and IFI meant that Model 2 had a better fit to the data. A lower chi-square value for this model denoted a greater explanation of model variance. Furthermore, the RMSEA values showed that Model 2 supported the criteria for an acceptable two-factor CFA model. Besides this, all items in Model 2 displayed substantial factor saturation, as indicated by their high factor loadings which is shown in Figure 1 (all \( \lambda \) coefficients were between .47-.70; all \( p < 0.00 \)).

<table>
<thead>
<tr>
<th>Models</th>
<th>Goodness-of-fit indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2 ) (df)</td>
<td>GFI</td>
</tr>
<tr>
<td>Model 1 (Original structure)</td>
<td>91.81(34)</td>
</tr>
<tr>
<td>Model 2 (EFA structure)</td>
<td>78.25(34)</td>
</tr>
</tbody>
</table>

**Note:** (Chi-square statistics; df: Degree of freedom for the likelihood ratio test of the model versus saturated; GFI: Goodness of Fit Index; AGFI: Adjusted Goodness of Fit Index; CFI: Comparative Fit Index; IFI: Incremental Fit Index; RMSEA: Root Mean Square Error of Approximation, AIC: Akaike’s Information Criterion.

In order to determine whether student gender affected the measurement model, the sample was split into groups of males and females. Thus, factorial invariance was assessed across gender by conducting confirmatory factor analyses using the best-fit model (Model 2), in which factor loadings and intercepts were free and then constrained. The test of configural invariance was supported in this case, since the same unconstrained factor structure which was simultaneously fit for the split groups yielded a good fit (CMIN/DF=2.40, CFI=0.86, RMSEA=0.076). Subsequently, metric invariance was examined by constraining the model (equal factor loadings across groups), and comparing its \( \chi^2 \) value with the unrestricted model. Results revealed a non-significant increase in the \( \chi^2 \) value in the constrained model relative to the unconstrained model, thereby establishing metric invariance. These findings indicated that the two groups showed invariance with respect to the factor structure.

**Reliability**

The present study also demonstrated that Model 2 had good reliability with CR value of 0.88, which was above the acceptable benchmark value of 0.70.

**Discussion**

Given the technological complexity of modern society, and the manner in which tests and evaluative situations determine the lives of people who take them, it comes as no surprise that the
testing situation evokes anxiety reactions among examinees[29]. Indeed, test anxiety figures prominently as one of the key variables in the ongoing drama of educational testing, as a source of both scholastic underachievement and psychological distress[29]. In fact, this appraisal has also been corroborated by a recent meta-analytic review which clearly indicated that test anxiety was significantly and negatively related to a wide range of educational performance outcomes, including standardized test scores, university entrance exam, and grade point average[30]. Therefore, timely detection of test anxiety is of utmost importance, especially amongst undergraduate students. Several scales have been developed over the years in order to identify test anxiety. The Westside Test Anxiety Scale (WTAS) is a widely used instrument, which was specifically designed to identify students with anxiety impairments, who could then benefit from an anxiety reduction intervention[20]. In the initial development and validation study of the WTAS, that involved two diverse groups of students, Driscoll (2007) reported sufficiently high validity[21]. However, utilization of WTAS has been comparatively limited in Malaysia. To our knowledge, this is the first study to examine the reliability and underlying factor structure of the WTAS in the Malaysian population.

The results of the exploratory factor analysis, in the data collected from this student sample, identified a two-factor model. This was somewhat similar to two of the categories which was suggested by the WTAS author, and originally designated as ‘Incapacity’ (items 1, 4, 5, 6, 8 and 10) and ‘Worry’ (items 2, 3, 7 and 9). However, despite the fact that a two-factor solution was replicated in the EFA, differences were observed between the two structures, regarding the distribution of items as well as number of items. Therefore, the two factors obtained from EFA were alternatively labelled as ‘Cognitive dysfunction’ (items 1, 3, 4, 7 and 8) and ‘Agony’ (items 2, 5, 6, 9 and 10). In addition, the ensuing confirmatory analysis also upheld the results of the exploratory analysis. It is worth noting that, even though the original two-factor model showed good fit indices, the revised two-factor model, guided by the results from EFA produced a significant improvement (P < 0.001) in model fit, and was consequently preferred as the best factor solution. In view of this finding, it could be recommended that the arrangement of a few items may be revised or adjusted, so as to enhance the discrimination among the two WTAS factors. This was furthermore supported by the assessment of Composite Reliability values for the two-factor EFA solution, which were also found to be relatively high. In order to further evaluate the psychometric soundness of the test scores, a comparative analysis of the similarity of the factors across gender (i.e. factorial invariance), was also conducted. Results of the multi-group CFA in this study revealed that the observed factor structure of the WTAS appeared to be equivalent in this setting, since the two-factor EFA model had good fit for both males and females. Thus, confirmation of gender invariance provided additional support for the validity of the WTAS as a self-report screening instrument, because it indicated that the measurement model of the latent test anxiety construct was comparable across gender. The present study represented a significant step in assessing and establishing the psychometric properties of the WTAS. However, there were some shortcomings which should be considered while interpreting the results. A convenience sampling technique was employed to recruit students for this study, and comprised of volunteers from only one faculty of the university. This may have limited the generalizability of the findings. Also, due to the cross-sectional nature of study design, it was difficult to derive causal relationships from the analysis. Another limitation associated with the present study was the utilization of a self-reported questionnaire, which may have potentially led to response bias. Therefore, future studies could include data collection by random sampling of students from various other faculties, as well as other universities or educational institutions. Further research on test anxiety could also include longitudinal studies in order to analyze causality.

Conclusion

The present study provided important evidence regarding the construct validity and multi-group factorial invariance of the WTAS in an educational setting in Malaysia. Even though the results of CFA showed that both models exhibited good fit indices, the two-factor EFA model obtained from this study presented the best factor solution. These factors were best interpreted using the recommended subscale designations of ‘Cognitive dysfunction’ and ‘Agony’. Additionally, this model also demonstrated gender invariance, which implied that differences in test scores observed between males and females truly reflected differences in test anxiety, rather than an artefact of the measurement method. As such, these data support the estimation that the WTAS is suitable for use as a screening tool for test anxiety among students in Malaysia.

References