DOCUMENT RESUME

ED 398 278 TM 025 481

AUTHOR	Nasser, Fadia; Takahashi, Tomone
TITLE	An Application of Confirmatory Factor Analysis with Item Parcels for Testing the Structure of Test Anxiety among Israeli-Arab High School Students.
PUB DATE	Apr 96
NOTE	37p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Arabic; *Factor Structure; Foreign Countries; High Schools; *High School Students; *Mathematics; *Sex Differences; Student Attitudes; *Test Anxiety; Test Items
IDENTIFIERS	*Confirmatory Factor Analysis; Israeli Arabs; *Item Parcels

ABSTRACT

The structure and the levels of test anxiety among Israeli-Arab high school students were examined using the Arabic version of I. G. Sarason's (1984) Reactions to Tests scale. The questionnaire was administered before a math examination to 226 female and 195 male students. The results of confirmatory factor analyses using eight item parcels consisting of three items each indicated that the four-factor model of Sarason fit the data best for both male and female students. Multiple group confirmatory factor analysis revealed that the number of factors, factor loading, and item residuals were invariant across gender. Latent mean analysis showed that girls reported higher test anxiety levels than boys in "worry," "tension," and "bodily symptoms," but not in "test irrelevant thinking." (Contains 1 figure, 5 tables, and 36 references.) (Author/SLD)

Running head: The Structure of Test Anxiety in Arab Population

An Application of Confirmatory Factor Analysis with Item Parcels for Testing the Structure of Test Anxitey among Israeli-Arab High School Students

Fadia Nasser

Tomone Takahashi[1]

University of Georgia

Paper presented at the annual meeting of National Council on Measurement in Education, April, 9 - 11, 1996, New York.

[1]The authors would like to thank Dr. J. Benson and Dr. J. Wisenbaker, for their valuable comments on earlier versions of this paper.

Address Correspondence to: Fadia Nasser, 325 Aderhold Hall, University of Georgia, Athens, GA, 30602, USA

Abstract

The structure and the levels of test anxiety among Israeli-Arab high school students were examined using the Arabic version of Sarason's Reactions to Tests scale. The questionnaire was administered before a math exam to 226 female and 195 male students. The results of confirmatory factor analyses using eight item parcels consisting of three items each indicated that the four-factor model of Sarason fit the data "best" for both male and female students. Multiple group confirmatory factor analysis revealed that the number of factors, factor loadings, and item residuals were invariant across gender. Latent mean analysis showed that girls reported higher test anxiety levels in "worry", "tension", "bodily symptoms", but not in "test irrelevant thinking".

KEY WORDS: Test anxiety structure, Reaction to Tests Scale, Confirmatory factor analysis, Item parcel, Gender differences

In our competitive society, tests are powerful tools widely used for decision-making. Individuals of all ages are frequently evaluated with respect to their performance, achievements, and abilities. Consequently, test anxiety has become one of the most frequently investigated constructs linked to under-achievement. Furthermore, test anxiety has been shown to affect students' performance and ability to profit from instruction (Tobias, 1980).

Dimensionality of Test Anxiety

It has long been theorized that the construct of test anxiety is multidimensional. Liebert and Morris (1967) initially proposed worry and emotionality components and this conceptualization of test anxiety was supported by several researchers (Morris, Davis & Hutchings, 1981; Spielberger,1980). The worry component embodies the cognitive aspect while the emotionality component taps one's self-reported physical reactions experienced by students during the testing situation. Tyron (1980) and Wine (1982) have argued that test anxiety should be viewed as including cognitive, emotional, behavioral and bodily reactions as elements of the construct. In addition to worry and emotionality, highly-anxious students experience bodily symptoms and direct their attention during tests to thoughts irrelevant to the task at hand.

Sarason (1984) cited many experimental studies related to cognitive interference found among high test-anxious students. Based on these findings he developed the Reactions to Tests (RTT) scale to measure these additional dimensions of test anxiety. The RTT scale consists of four subscales which are labeled "tension", "worry", "test irrelevant thinking", and "bodily symptoms". The

range of the correlations among the four subscales as reported by Sarason is .24-.69. This range with a similar pattern of factor correlation was also reported by Flett, Blankstein, and Boase (1987).

Test Anxiety Studies in Arab Populations

Judging from the literature, test anxiety is a universal phenomena (El-Zahhar & Hocevar, 1991). However, most of the findings about test anxiety measured either by Sarason's RTT or other test anxiety scales (Morris, Davis, & Hutchings, 1981; Spielberger, 1980) are based on Western samples. Most of the studies of test anxiety have emphasized the relationship between test anxiety and performance and/or gender difference on levels of test anxiety. Furthermore, validation of the RTT as conceptualized by Sarason (1984) is limited to only a few studies (e.g., Benson & Bandalos, 1992).

In the last decade a few cross-cultural studies of test anxiety involving Arab populations were conducted (Benson & El-Zahhar, 1994; Hocevar & El-Zahhar, 1988). These studies mostly focused on the levels of test anxiety across cultures and gender. With regard to gender differences on the levels of test anxiety, the findings of studies of test anxiety in Arab populations are consistent with those from American populations indicating that levels of anxiety are higher among females compared with males. However, levels of anxiety in the Arab populations were found to be higher when compared with American populations (Ahlawat, 1989; Benson & El-Zahhar, 1994; El-Zahhar & Hocevar, 1991). The higher level of anxiety among Arab students was interpreted as a

consequence of the extreme importance of the test for the high school students in their society (El-Zahhar, 1991).

Investigating test anxiety in the Arab population is limited to a few studies, and much less has been done on the Israeli-Arab population, which will be one of the foci of this study. A previous study of test anxiety and test performance in the Israeli-Arab population reveals results that are consistent with the findings that exist in the literature with regard to gender differences in the levels of test anxiety (Birenbaum & Nasser, 1994). However, no studies were found in the literature that confirm the structure of test anxiety in the Israeli-Arab population as measured by the RTT scale.

Applications of Confirmatory Factor Analysis to Test Anxiety Studies

With the widespread use of LISREL, investigating the structure of psychological constructs by using structural equation modeling (SEM) has attracted more researchers, and more findings are being compiled in the literature. Test anxiety is not an exception in this regard. A confirmatory factor analysis (CFA) based on the measurement model proposed by Jöreskog (1969) can be used to examine the factor structure of latent variables. A model is hypothesized based on theory and the maximum likelihood estimation method is usually used to calculate the parameter estimates based on the hypothesized model. The fit of the model is estimated by the model's ability to reproduce the covariance matrix of the observed variables. Among the few published studies of test anxiety in which CFA was used are studies done by Benson and Bandalos (1992),

Benson, Bandalos and Hutchinson (1994), and Benson and Tippets (1990). The studies by Benson and her colleagues involved the structure of test anxiety, structure invariance across gender and cross validation of the results. Furthermore, Hocevar and Chiou (1995) suggested that CFA is the most efficient method for cross-cultural validation of personality constructs including test anxiety.

The Use of Item parcels in SEM Studies

Maximum likelihood estimation is sensitive both to the number of observations and to the number of parameters to be estimated (Anderson and Gerbing, 1984). This estimation method is based on the assumption that the data are continuous and normally distributed. However, this assumption is frequently violated in CFA, especially when categorical variables are analyzed, and can result in misleading findings and conclusions about the factor structure under study (Bernstein & Teng, 1989).

In the published literature involving the use of SEM to study latent variables, researchers have summed individual items to create item parcels. These item parcels are then used as the observed variables in the model of interest. Item parcels have been formed to simplify the models by creating smaller numbers of observed variables and to create indicators of the latent constructs which are more like continuous variables. Another advantage of summing items and forming item parcels involves creating more continuous variables, which allows for distributions closer to normal. Although the information from variances and covariances of the individual items will be lost, the item parcels

are more likely to meet the assumptions of maximum likelihood estimation.

Some researchers have created item parcels by forming random combinations of items such as split halves or split thirds or odd-even, depending on the number of items and the needs of the model (Prats, 1990). Other researchers have created item parcels based on size of the item parcel means, standard diviations, and skew (Schau, Stevens, Dauphinee, & Vecchio, 1995). It is obvious that the above considerations in forming the item parcels are purely statistical and ignore theory and the content similarity of the items. Furthermore, such random combinations of items provide inconsistent results in terms of model fit (Prats, 1990).

Purpose of the Study

The purpose of this study is (a) to examine the factor structure of test anxiety in Israeli-Arab high school students as measured with the Arabic version of Sarason's RTT scale; (b) to examine whether the factor structure and latent means of test anxiety are equal across gender. To accomplish these purposes, we used item parcels rather than items as measurement variables of the factors of test anxiety.

Method

Sample

The sample consisted of 421 tenth graders (ages 15-16) from 15 classes of two Arab high schools (216 from one school and 205 from the other school) in the central district of Israel.
Of the participants, 195 were boys and 226 were girls. All students in the two schools were Muslims. The two schools are among the

largest schools in the Arab sector in Israel. The mean socioeconomic status (SES) of the students' families in each school is very close to the national SES mean (Nasser, 1989).

Instrument

The RTT scale which was developed by Sarason (1984) consists of 40 items, four-point likert rating scale. The Arabic version of the RTT questionnaire was a translation of the Hebrew version developed by Birenbaum and Montag (1986). The translation to the Arabic version was done by the first author of this study and the back translation was done by a university professor who is bilingual in Hebrew and Arabic.

Cronbach's Alpha coefficients for the RTT scale as reported by Sarason (1984) were .78 for the total scale and .68 to .81 for the subscales. Cronbach's Alpha coefficients for the Arabic RTT total scale is .94 for the current sample, and .93 for boys and girls separately. The reliablities of the subscales for the entire sample ranged from .80 to .87. For boys, they range .77 to .81, and for girls, .81 to .86.

Procedure

The RTT was administered to the participants before a mathematics exam during their regular class sessions by the first author. The mathematics test was a scheduled test for the topic and students prepared for it the way they did for other mathematics tests. The participants and their parents were told that the purpose of the study was to gain better understanding of the relationship between the test performance and test anxiety in order to design an intervention program to benefit those who need help

coping with test anxiety. The participants were assured that their responses to the RTT questionnaire would not be released to the school authorities without their consent, and that they only would be used for research purposes.

Data Analysis

At the first stage, exploratory factor analysis (EFA) and CFA with 40 items were conducted to test the fit of the original scale and to provide supplementary information used for developing item parcels.

The goal of the second stage was to create item parcels consisting of three items each. Items belonging to the same subscale (10 items) according to Sarason's theory were examined in terms of the similarity of their content. We also consulted with the results of exploratory factor analysis (EFA). We grouped items which were similar in content and loaded at least .30 on the same factor. When more than three items fulfilled the two criteria stated above, we chose to group three items which had the highest item-factor correlations. We decided to use three items rather than two to form each item parcel, because this combination better meets the continuity assumption. Using more than three items per item parcel would lead to not having at least two indicators per latent variable, which would create a partial identification problem. Only 24 of 40 items met the criteria. Thus, eight item parcels were used in the analyses.

The 24 items from which the eight item parcels were formed are shown in Appendix 1. Once the reduced set of items was determined, a CFA with the 24 items was conducted based on the hypothesized

four-factor structure to: (a) study the model-data fit of the 24-item model compared with the original 40-item model, (b) study the model-data fit of the 24-item model compared with the eight item parcels model, and (c) obtain item level reliability coefficients.

In the next step, four alternative models which are based on competing theories in the measurement of test anxiety were specified a priori (Figure 1). The specified models are:

Model 1: Four-factor model. This model is based on Sarason's theory which hypothesized test anxiety as an four dimensional structure (worry, test irrelevant thinking, tension, and bodily symptoms). In this model, each of the four factors is measured by two item parcels, and satisfied the necessary condition of identification (t-rule). However, to satisfy the sufficient rule of identification, the factors must be correlated (Bollen, 1989). This condition is assumed to be satisfied based on the findings of the previous research (Sarason, 1984).

Model 2: Two-factor model (a). Since both worry and test irrelevant thinking are two cognitive aspects of test anxiety, they were grouped to form one dimension. Also bodily symptoms and tension were grouped to form the second factor because both of are reflections of emotional reactions to testing situations.

Model 3: Two-factor model (b). Since test irrelevant thinking is a new concept proposed by Sarason, the two item parcels which define the subscale of "test irrelevant thinking" were omitted to test the fit of the theoretical two factor structure as proposed by Spielberger. In this model 'tension' and 'bodily symptoms' were collapsed into one factor which is called 'emotionality'

(Birenbaum, & Nasser, 1994). Worry, which was represented by the two worry item parcels, made up the second factor.

Model 4: Three-factor model. Test irrelevant thinking is brought back and is treated as a separate factor. Tension and bodily symptoms are assumed to be indicators of an emotionality factor, and worry made up the third factor.

Models 2 and 3 are alternative models which represent variations of Spielberger's two-factor model of test anxiety. They were proposed as representatives of rival theories to Sarason's four-factor theory of test anxiety. Model 4 joins aspects of Sarason and Spielberger's conceptualizations of test anxiety.

A CFA was conducted on the four models for males and females separately using the item parcels as indicators. Separate covariance matrices for boys and girls were used as input to the LISREL VII program (Jöreskog & Sörbom, 1988) to analyze the models of interest in this study. All the CFA results were obtained with maximum likelihood estimation. Model fit was evaluated in terms of acceptable criteria for indices of fit, parsimony, and meaningfulness. The fit indices were selected both from absolute and incremental indices of fit based on Hoyle and Panter's (1995) recommendation. The fit indices used in this study include: three absolute indices, chi-squared to degrees of freedom ratio, the Goodness of Fit Index (GFI), and the root mean square residuals (RMR); and two incremental fit indices, the Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI). The last two indices seem to be less influenced by sample size compared with other fit indices (Hu & Bentler, 1995; Hoyle & Panter, 1995). Judging from the

literature, the acceptable evaluation criteria for the listed indices are: chi-squared-to-degrees-of freedom-ratio should not exceed 2.00; the other indices of fit should exceed .90. The CALIS procedure in SAS 6.04 was also used to obtain the following fit indices: the Tucker-Lewis Index (TLI), the comparative fit index (CFI), and expected cross-validation index (ECVI) along with its confidence interval.

With regard to cross-validation, the replicability of the alternative models to other samples from the same population is tested by estimating the expected cross-validation index (ECVI) for a single sample as proposed by Browne and Cudeck (1989, 1993). ECVI reflects the expected overall discrepancy over all possible calibration samples. Smaller values of ECVI indicate a higher probability that the model will be replicable across samples from the same population.

The four models will be tested and evaluated for girls and boys separately, and the "best model" for each group will be selected. If the selected model in the previous process is the same for boys and girls, multiple group CFA will be considered to test model invariance for boys and girls. The invariance analyses will be used to test the following hypotheses:

(A) Are the observed variance-covariance matrices equal for boys and girls?

(B) Are the numbers of factors equal for boys and girls?

(C) Are the factor loadings equal for boys and girls?

(D) Are the variances and covariances among factors equal for boys and girls?

(E) Are the residuals of the observed variables equal for boys and girls?

Most of the literature about test anxiety involves the comparison of levels of test anxiety across gender and/or the relation between test anxiety and performance in evaluative situations. These studies compare levels of test anxiety by comparing the means of observed variables without taking measurement error explicitly into account. Because it is known that similar observed variable means do not necessarily lead to similar latent variable means, in the current study, we attempt to answer the question of whether the latent means of the factors are equal for boys and girls.

Results

Preliminary Analysis

With regard to the four-factor model with 40 items, even though the values of χ^2/df were less than 2.0, the values of the fit indices showed that the model did not fit the data well. (χ^2=1217.59, df=734, p=.000, χ^2/df=1.66, GFI=.76, CFI=.80, TLI=.79, RMR=.05 for boys, and χ^2=1259.27 df=734, p=.000, χ^2/df=1.72, GFI=.79, CFI=.84, TLI=.83, RMR=.06 for girls). Judging from the values of the fit indices, CFA with the 24 selected items showed better fit than the 40-item model, but still the fit was not at a satisfactory level (χ^2=420.56, df=246, p=.000, χ^2/df=1.71, GFI=.85, CFI=.88, TLI=.86, RMR=.05 for boys, and χ^2=372.70, df=246, p=.000, χ^2/df=1.52, GFI=.88, CFI=.93, TLI=.92, RMR=.06 for girls). The remaining analyses were conducted with eight item parcels each consisting of three items.

Descriptive Statistics

Descriptive statistics of the eight item parcels are summarized in Table 1. Univariate non-normality was a problem for two of the subscales: Test Irrelevant Thinking-1 for boys (kurtosis 2.15 and skewness 1.57) and girls (kurtosis 3.35 and skewness 1.85) and Bodily Symptoms-1 for boys only (kurtosis 3.21 and skewness 1.84). Mardia's measure of multivariate kurtosis indicates that using item parcels rather than individual items improve multivariate normality (161.18 vs. 20.43 for boys, and 72.52 vs. 2.90 for girls).

Insert Table 1 about here

Model Comparisons

The model-data fit of the four-factor model (model 1) based on Sarason's theory was compared with three alternative models (models 2 to 4) for boys and girls separately. The four models cannot be compared statistically, because they are not nested in each other. Of the four models tested, only Model 1 showed a p value greater than .05 (.14 for boys and .33 for girls) along with the smallest χ^2/df ratio (1.40 for boys and 1.13 for girls), and the highest ad hoc fit indices (> .90) for both boys and girls (see Table 2). Thus, we concluded that the four-factor model based upon Sarason's theory and scale of test anxiety using eight item parcels fits the data better than the other three models. Model 1 also turned out to fit the data better than the four-factor models based on individual items of 40 and 24.

Insert Table 2 about here

Four-Factor Model with Eight Item parcels (Model 1)

Item level reliabilities as indicated by squared multiple correlation based on the 24 individual items that made up the item parcels ranged from .20 to .61 for boys and from .17 to .62 for girls (see Appendix 1). Item parcel level reliabilities are higher than those of individual items, and they ranged from .44 to .70 for boys and from .46 to .78 for girls.

Item parcel factor loadings are also generally higher than those of the 24 individual items and ranged from .66 to .89 (see Table 3). These results indicated that item parcels are more reliable and perhaps better indicators than individual items.

Insert Table 3 about here

Factor correlations are presented in Table 4. Correlations among the factors are fairly high (.65 to .92 for boys, and .48 to .95 for girls), in particular, the correlation between the worry and tension factors, which exceeds .90. The high correlation suggests that these factors are almost identical. When these two factors were collapsed, the fit of the resulting three-factor model was as good as the four-factor model (χ^2=26.00, df=17, p=.075, χ^2/df=1.53, GFI=.97, CFI=.98, TLI=.97 for boys, and χ^2=26.22, df=17, p=.071, χ^2/df=1.54, GFI=.97, CFI=.99, TLI=.99 for girls).

Insert Table 4 about here

Cross-Validation

The CALIS procedure provided the ECVI along with its confidence interval for each model. The results in Table 2 indicate that model 1 has the smallest ECVI among the models with eight indicators, and that the 90% confidence interval associated with it includes zero. These results indicate that the discrepancy over all possible calibration samples would not differ statistically from the present results. Two of the three alternative models (models 2 and 4) which were specified a priori revealed larger ECVI, and the confidence intervals associated with the three alternative models did not include zero. Thus, these models would not likely cross-validate well across other samples drawn from the same population. The three-factor model, in which worry and tension were collapsed into one factor based on the high correlation between the two, yielded ECVI and confidence intervals [.34, (.00, .43) for boys and .29 (.00, .37) for girls] similar to model 1 (see Table 2). This indicates that this model would also cross-validate as well as model 1. Furthermore, the point estimate of ECVI corresponding to model 3 was the smallest for boys and girls. However, the confidence interval corresponding to this model did not include zero. The smallest point estimate of ECVI for model 3 might be attributed to smaller standard error, because it consists of less elements compared with the other models. Judging from the results, among the models tested, only the model with three factors in which

tension and worry are collapsed into one factor might be considered as a competing model with model 1.

Model Invariance across Gender

The findings that model 1 fit the data well for each of the groups separately does not ensure that it will fit the two groups when they are compared simultaneously. To examine the invariance of model 1 across gender, a series of multiple group CFAs were conducted.

Table 5 summarizes the findings of the multiple group comparisons of model 1. The results indicate that the observed variance-covariance matrices are not the same across gender (χ^2=58.50, df=16, p=.010).

Insert Table 5 about here

A sequence of hypotheses testing the addition of equalities across gender was used to pinpoint how the observed covariance matrices differed for males and females. As indicated by the chi-square differences in the bottom part of Table 5, the results of the model comparisons revealed that the four-factor model resulted in an equal number of factors and equal loadings for boys and girls. However, the model was not invariant when the restrictions of equal factor variances and covariances were added.

The factor correlation matrices for boys and girls (Table 4) showed that the difference of the correlation between tension and test irrelevant thinking is the largest (.68 for boys and .48 for girls). The modification index for the factor variance-covariance

matrix also indicated that the covariance between tension and test irrelevant thinking was the largest (10.25) one. Thus, it appears that large difference of the covariance between tension and test irrelevant thinking for boys and girls is the element most responsible for the gender differences. This result was supported by a series of multiple group analyses. Nine of the 10 analyses in which the correlation between test irrelevant thinking and tension was constrained indicated that the difference between males and females was significant. Only the analysis in which that correlation was not constrained (Model D') resulted in no group difference (see Table 5).

Because the correlations between the factors were not equal across gender, it was not meaningful to impose the additional restriction of equal residuals for boys and girls. Therefore, the restriction of equal residuals was imposed on model C to test the equality of residuals beyond the equality of number of factors and factor loadings across gender (Model E, in Table 5). Testing the difference between models E and C, which are nested, indicated that the residuals were invariant across gender ($\Delta\chi^2$=14.23, df=13, p = .36).

Latent Means

To compare the levels of test anxiety across gender, latent means were introduced into the model. This was done in two steps. First, the four-factor model for boys and girls was tested by adding the restriction of equality of the observed means to the invariant factor loadings and nine of 10 elements of the factor variance and covariance matrix (model F in Table 5). The results of

this analysis indicated that the means of the observed variables (item parcels) are invariant across gender (comparison model F to model D'). In the second step, we added the restriction of invariant latent means to the model F (model G in Table 5). The addition of the invariant latent means resulted in significant chi-square difference for the overall model, which indicated that latent means were not equal. However, examination of the t-values corresponding to each of the four latent means indicated that only the latent means of test irrelevant thinking were not significant ($t < 2.0$). The latent means on the other three factors were significantly different for boys and girls ($t > 2.0$). The results provide evidence that girls have higher latent means than boys on "worry", "tension", and "bodily symptoms", but not on "test irrelevant thinking".

Discussion

This study had two main objectives. The first of these was to test the structure of test anxiety of Israeli-Arab high school students and to see whether the four factor structure proposed by Sarason could be extended to this sample. The findings indicated that the four factor structure with 40 items did not fit the data well. This finding is consistent with the results reported by Benson and Bandalos (1992) for American college students. Several reasons may be responsible for the misfit. One reason might be violation of one or more of the assumptions underlying maximum likelihood estimation method such as, the need for continuous variables and normal distribution. Another reason that may be responsible for the misfit may also be the low reliabilities of the

measurement variables. Still another reason might be model misspecification. Because of the four-factor theory of Sarason, we formed item parcels to meet the statistical assumptions of CFA closely, instead of considering alternative models using the 40 models based on statistical consideration alone.

The findings involving model 1 (four-factor/eight-item parcel) indicated that the model fits the data very well. Model 1 also fits the data much better than the model based on the same 24 individual items. When the item parcels are created based item content and EFA factor loading patterns, the indicators are more reliable and result in less specification errors.

To extend the structure from the first objective of the study, we compared model 1 with three alternative models. The finding that a single theoretically proposed model fits the data well is important, but this finding can be strengthened by comparing several theoretically plausible models (Jöreskog, 1993). Thus, three versions of Spielberger's two factor model were considered.

The results of overall fit as indicated by chi-square, chi-square to degrees of freedom ratio, and three fit indices (GFI, CFI, TLI) favored the four-factor structure. This means that the four-factor structure as incorporated in Sarason's RTT scale holds in this sample of Israeli-Arab high-school students. These findings were also supported by the cross-validation results.

In model 1, the correlations between the factors were moderate to high, especially the correlation between the worry and the tension factors, which was extremely high for both boys and girls. Except for the extremely high correlation between worry and

tension, the factor correlation pattern was consistent with previous studies (Benson & Bandalos, 1992; Sarason, 1984). The high correlation between these two factors implied that they might collapse into one factor and result in a three-factor model. Although this model has no theoretical support, it was considered, because there is no previous research on the structure of test anxiety in an Israeli-Arab population. The results involving this model revealed that the model fits the data almost as well as the four-factor model. It is safe to conclude, for the current sample with eight item parcels formed from 24 items, that the factor structure does not contradict the four-factor structure proposed by Sarason. However, we cannot ignore the possibility that an Israeli-Arab population may have a different factor structure, which suggests that worry and tension create one dimension of test anxiety. Students in this population experience different kinds of anxiety as a result of their socio-political situation and their status as a minority. It seems that they may not differentiate between worry and tension in a threatening situation. It is also worth mentioning that translation issues might be responsible for the lack of distinction between the two factors. There is a need for further research and replications of these findings to confirm that an alternative theoretical structure of test anxiety is necessary for this population.

The second objective of the study was to examine whether the factor structure and latent means of test anxiety were equivalent across gender. There is a consensus in the literature that girls report higher levels of test anxiety compared with boys. Several

theories offered an explanation for the different levels of test anxiety for boys and girls. One of these theories suggests that boys are less likely to admit their feelings of anxiety compared with girls (Maccoby & Jacklin, 1974).

Another theory suggests that males and females are socialized to experience and to respond to evaluative situations differently (Arch, 1987). The difference could also be explained by differences in the factor structure of test anxiety across gender. With regard to the factor structure, when model 1 (four factors, eight item parcels) was compared across gender by multiple group analysis, the results showed that all the measurement parameters and most of the structural parameters were invariant. The only structural parameter to differ was between tension and test irrelevant thinking. It seems that the test irrelevant thinking and tension dimensions of test anxiety are less distinct for boys than for girls. This difference might be interpreted as another aspect of gender differences in responding to test anxiety. Girls admit negative feelings more than boys do.

Most researchers have reported gender differences in the levels of test anxiety based on observed measures. In the current study the gender differences were examined by comparing latent means. The importance of studying differences in latent means compared with observed means is that latent means are free from measurement errors (specific factors and random measurement errors). Generally the latent means for boys and girls were different, which is consistent with the findings of the previous research. However, the latent means for test irrelevant thinking

were similar for boys and girls. This finding is consistent with those of previous research, which indicated that the large difference in levels of anxiety lies in the worry and emotional aspects, especially the emotional aspects (tension and bodily symptoms; Sarason, 1984; El-Zahhar & Hocevar, 1991). The similar low levels of test irrelevant thinking among boys and girls implies that this factor is somewhat different from the other three factors of test anxiety. The relatively low correlations between test irrelevant thinking and the other subscales also questions the relevance of this cognitive interference component introduced by Sarason (1984) to test anxiety. Certainly, this conjecture needs to be supported by further research.

The present study is the first to test the structure of test anxiety in the Israeli-Arab population. Furthermore, the RTT scale was administered prior to a mathematics test. These two facts imposed several limitations on the results of this study. Among these limitations, the generalizability of the results is limited to the sample tested in the study. The replicability of the "best" model will be conditioned on the similarity of future samples to the sample involved in this study. Another limitation stems from the contexts in which the RTT was administered. It might be that the structure tested in this study is more likely to be the structure of state test anxiety or mathematics anxiety, rather than trait test anxiety. Therefore, future research is needed to test the structure of test anxiety and the stability of the structure in different contexts.

References

Ahlawat, K. S. (1989). Psychometric properties of the Yarmouk test anxiety inventory (Y-TAI). In R, Schwarzer, H. M., van der Ploeg., & C. D. Spielberger (Eds.), Advances in test anxiety research (Vol. 6, pp. 263-278). Lisse/Hilsdale, NJ: Swets & Zeitlinger/Erlbaum.

Anderson, J. C., & Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solutions and confirmatory factor analysis. Psychometrika, 45, 155-173.

Arch, E. C. (1987). Differential responses of females and males to evaluative stress: Anxiety, self-esteem, efficacy, and willingness to participate. In R, Schwarzer, H. M., van der Ploeg., & C. D.Spielperger (Eds.), Advances in test anxiety research (Vol. 5, pp. 97-106). Lisse/Hilsdale, NJ: Swets & Zeitlinger/Erlbaum.

Benson, J., & Bandalos, D. (1992). Second-order confirmatory factor analysis of reactions to tests scale with cross-validation. Multivariate Behavioral Research, 27, 459-487.

Benson, J., Bandalos, D., & Hutchinson, S. (1994). Modeling test anxiety among men and women. Anxiety, Stress and Coping, 7, 131-148.

Benson, J., & El-Zahhar, N (1994). Further refinement and validation of the revised test anxiety scale. Structural Equation Modeling, 1, 203-221.

Benson, J., & Tippets, E. (1990). Confirmatory factor analysis of the test anxiety inventory. In C. Spielberger, & R. Diaz-Guerrero (Eds.), Cross cultural anxiety (Vol. 4 pp. 149-156). New York: Hemisphere/Taylor-Francis Publishing.

Bernstein, H., & Teng. G. (1989). Factoring items and factoring scales are different: spurious evidence for multidimensionality due to item categorization. Psychological Bulletin, 105, 467-477.

Birenbaum, M., & Montag, I. (1986). Reactions to tests: A Hebrew version of Sarason's RTT Questionnaire. Unpublished manuscript, Tel Aviv University, School of Education, Unit of Measurement and Evaluation Research, Israel (In Hebrew).

Birenbaum, M., & Nasser, F. (1994). On the relationship between test anxiety and test performance. Measurement and Evaluation in Counseling and Development, 27, 293-301.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. Multivariate Behavioral Research, 24, 445-455.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. A., & Long, J. S. (Eds.), Testing structural equation models (pp. 136-162). Newbury Park, CA: Sage.

El-Zahhar, N., & Hocevar, D. (1991). Cultural and sexual differences in test anxiety, trait anxiety and arousability, Egypt, Brazil, and the United States. Journal of Cross-Cultural Psychology, 22, 238-249.

Flett, G. L., Blankstein, K. R., & Boase, P. (1987). Self-focused attention in test anxiety and depression. Journal of Social Behavior and Personality, 2, 259-266.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. Journal for Research in Mathematics Education, 21, 33-46.

Hocevar, D., & Chiou, H. (1995, April). Methodological approaches and issues in cross-cultural psychometrics: The covariance modeling perspective. Paper presented at the 1995 conference of the National Council on Measurement in Education, San Francisco, CA.

Hocevar, D., & El-Zahhar, N. (1988). Arousability, trait anxiety and worry and emotionality components of test anxiety. Anxiety Research, 1, 99-113.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 158-176). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1995). Evaluation of model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 76-99). Thousand Oaks, CA: Sage.

Jöreskog, K. (1969). A general approach to a confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-202.

Jöreskog, K. (1993). Testing structural equation models. In Bollen, K. A., & Long, J. S. (Eds.), Testing structural equation models (pp. 294-316). Newbury Park, CA: Sage.

Jöreskog, K., & Sörbom, D. (1988). LISREL, VII: A guide to the program and application. Mooresville, IN: Scientific Software.

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some

initial data. Psychological Reports, 20, 975-978.

Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.

Morris, L. W., Kellaway, D. S., & Smith, D. H. (1978). Mathematics anxiety rating scale: Predicting anxiety experiences and academic performance in two groups of students. Journal of Educational Psychology, 70, 589-594.

Morris, L. W., Davis, M., & Hutchings, C. J. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. Journal of Educational Psychology, 73, 541-555.

Nasser, F. (1989). Effects of sex, test anxiety and item sequence on performance on diagnostic test in exponents. Unpublished master's thesis. Tel Aviv University, Israel (In Hebrew).

Prats, D. C. (1990, April). The effects of forming item parcels on the construct validity of the test anxiety inventory. Paper presented at the National Council on Measurement in Education, Boston.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. Journal of Personality and Social Psychology, 46, 929-938.

Schau, C., Stevens, J., Dauphinee, T., & Vecchio, A. D. (1995). The development and validation of the survey of attitudes toward statistics. Educational and Psychological Measurement, 55, 868-875.

Speilberger, C. D. (1980). Preliminary professional manual for

the Test Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Tobias, S. (1985). Anxiety and instruction. In I. G. Sarason (Ed.), Test anxiety: Theory, research, and applications (pp.289-310). Hillsdale, NJ: Erlbaum.

Tyron, G. (1980). The measurement and treatment of test anxiety. Review of Educational Research, 50, 343-372.

Wighfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. Journal of Educational Psychology, 80, 210-216.

Wine, J. D. (1982). Evaluation anxiety: A cognitive, attentional construct. In H. W. Krohne & L. Laux (Eds.), Achievement, stress, and anxiety (pp. 207-219). Washington, DC: Hemisphere.

Appendix 1

Reliability of 24 items used for 8 parcels

Item Reliability		
Boys	Girls	Items
Worry-1		
.56	.38	While taking a test, I often think about how difficult it is.
.37	.50	Thoughts of doing poorly interfere with my concentration during tests.
.32	.53	During test, I think about how poorly I am doing.
Worry-2		
.20	.18	The thought, " what happens if I fail this test?" goes through my mind during the test.
.29	.27	During difficult test, I worry whether I will pass it.
.41	.40	Before taking a test, I worry about failure.
Tension-1		
.44	.44	While taking a test I feel tense.
.34	.41	I find myself become anxious the day of the test.
.44	.48	I am anxious about tests.
Tension-2		
.31	.20	I wish tests did not bother me so much.
.37	.50	I feel panicky during tests.
.35	.49	I have an uneasy feeling before an important test.
Test Irrelevant Thinking-1		
.34	.34	During tests I find myself thinking of things unrelated to the material being tested.
.40	.43	Irrelevant bits of information pop into my head during a test.
.31	.45	I think about current events during a test.
Test Irrelevant Thinking-2		
.35	.34	My mind wanders during tests.
.22	.26	While taking a test, I often do not pay attention to the question.
.45	.54	I have fantasies a few times during a test.
Bodily Symptoms-1		
.30	.58	I get a headache during an important test.
.56	.35	I sometimes feel dizzy after a test.
.61	.62	I get a headache before a test.
Bodily Symptoms-2		
.32	.27	I become aware of my body during tests (feeling itches, pain, sweat, nausea)
.31	.17	My hands often feel cold before and during a test.
.27	.30	I sometimes find myself trembling before or during tests.

Table 1

Descriptive Statistics and correlation matrices of the 8 parcels for Boys and Girls

Mini-scales	Mean	SD	Kurto-sis	Skew-ness	correlations W-1	W-2	TEN-1	TEN-2	TIT-1	TIT-2	BS-1	
Boys												
1. Worry-1	6.42	2.18	-0.33	0.45								W-1
2. Worry-2	6.94	2.23	-0.58	0.15	.577							W-2
3. Tension-1	6.34	2.18	-0.11	0.65	.573	.519						TEN-1
4. Tension-2	6.04	2.17	-0.44	0.45	.624	.527	.641					TEN-2
5. Test Irrelevant Thinking-1	4.31	1.73	2.15	1.57	.402	.228	.363	.367				TIT-1
6. Test Irrelevant Thinking-2	4.87	1.84	0.51	1.01	.389	.329	.439	.469	.554			TIT-2
7. Bodily Symptoms-1	4.34	1.91	3.21	1.84	.332	.382	.393	.434	.294	.401		BS-1
8. Bodily Symptoms-2	4.81	1.86	1.84	1.32	.425	.362	.508	.490	.230	.362	.554	BS-2
Girls												
1. Worry-1	7.35	2.40	-0.78	0.26								W-1
2. Worry-2	8.03	2.30	-0.83	-0.07	.533							W-2
3. Tension-1	8.12	2.48	-1.04	0.01	.606	.588						TEN-1
4. Tension-2	7.41	2.48	-0.96	0.20	.607	.533	.690					TEN-2
5. Test Irrelevant Thinking-1	4.35	1.92	3.35	1.85	.370	.259	.297	.287				TIT-1
6. Test Irrelevant Thinking-2	5.00	2.09	0.68	1.10	.473	.313	.370	.329	.629			TIT-2
7. Bodily Symptoms-1	5.59	2.47	0.03	0.92	.414	.339	.516	.510	.300	.330		BS-1
8. Bodily Symptoms-2	6.02	2.17	-0.02	0.67	.501	.371	.512	.466	.323	.430	.559	BS-2

Table 2

Model Comparison

Model	χ^2	df	p	χ^2/df	GFI	CFI	TLI	ECVI (90% CI)
Boys								
1. 4 factors/8 indicators	19.56	14	.14	1.40	.98	.99	.98	.34 (.00, .42)
2. 2 factors/8 indicators	79.40	19	.001	4.18	.91	.90	.85	.59 (.47, .76)
3. 2 factors/6 indicators	29.69	8	.001	3.71	.95	.95	.91	.29 (.22, .40)
4. 3 factors/8 indicators	42.97	17	.001	2.53	.95	.96	.93	.43 (.35, .55)
Girls								
1. 4 factors/8 indicators	15.80	14	.33	1.13	.98	.99	.99	.27 (.00, .34)
2. 2 factors/8 indicators	111.74	19	.001	5.88	.89	.88	.82	.65 (.52, .82)
3. 2 factors/6 indicators	29.98	8	.001	3.75	.96	.96	.93	.25 (.19, .35)
4. 3 factors/8 indicators	44.91	17	.001	2.64	.95	.96	.94	.38 (.30, .48)

Note. GFI = Goodness-of-fit index; CFI = Comparative fit index; TLI = Tucker-Lewis index; ECVI = Expected cross-validation index; CI = Confidence interval.

Table 3

Factor Loadings of Four-factor Model for Boys and Girls

Factors	Indicators	Boys		Girls	
		Loadings	Reliability	Loadings	Reliability
Worry	W-1	.82	.66	.78	.61
	W-2	.71	.50	.68	.65
Tension	Ten-1	.78	.61	.85	.72
	Ten-2	.82	.67	.81	.66
Test Irrelevant	TIT-1	.66	.44	.71	.50
Thinking	TIT-2	.84	.70	.89	.78
Bodily Symptoms	BS-1	.70	.49	.73	.54
	BS-2	.79	.62	.76	.58

Note. All the loadings are significant ($p < .05$)

Table 4

Factor Correlation for Boys and Girls

	Worry	Tension	TIT
Boys			
Tension	.92		
TIT	.59	.68	
Bodily symptoms	.65	.77	.58
Girls			
Tension	.95		
TIT	.62	.48	
Bodily symptoms	.76	.80	.58

Table 5

Summary of the Multiple Group Analysis (Girls and Boys)

Model	χ^2	df	p	GFI Boys	GFI Girls
(A) Equal covariance matrices	58.50	36	.01	.95	.98
(B) Equal number of factors	35.36	28	.16	.98	.98
(C) Equal number of factors & Equal loadings	37.26	32	.24	.98	.98
(D) Equal number of factors, Equal loadings, & Equal factor variances and covariances	57.17	42	.06	.96	.97
(D') Equal number of factors, Equal loadings, & Equal factor variances and covariances (cov. between tension and test irrelevant thinking was not constrained)	46.50	41	.27	.97	.98
(E) Equal number of factors, Equal loadings, & Equal item residuals	51.49	40	.11	.97	.98
(F) Equal number of factors, Equal loadings, Equal factor variances and Covariances (cov. between tension and test irrelevant thinking was not constrained), & Equal observed means	52.24	45	.21	.97	.98
(G) Equal number of factors, Equal loadings, Equal factor variances and covariances (cov. between tension and test irrelevant thinking was not constrained), Equal observed means, & Equal latent means	134.39	49	.00	.97	.98

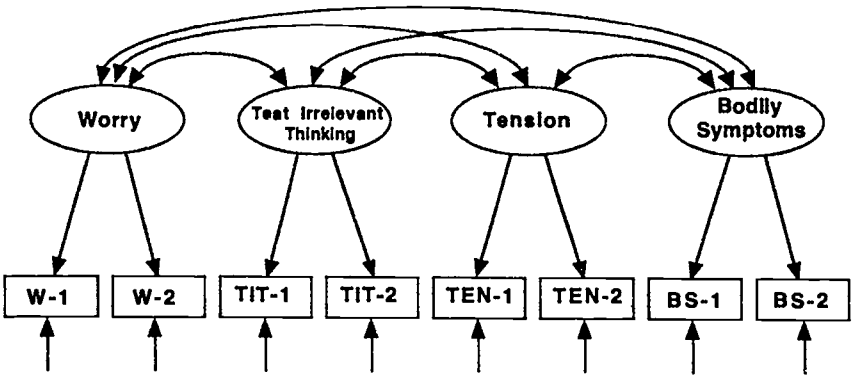
Model Comparison	$\Delta\chi^2$	Δdf	p
(C) - (B)	1.90	4	.75
(D) - (C)	19.91	10	.03
(D') - (C)	9.24	9	.42
(E) - (C)	14.23	8	.08
(G) - (F)	82.15	4	.00

Note. GFI = Goodness-of-fit index.

Sarason's Model

Model 1

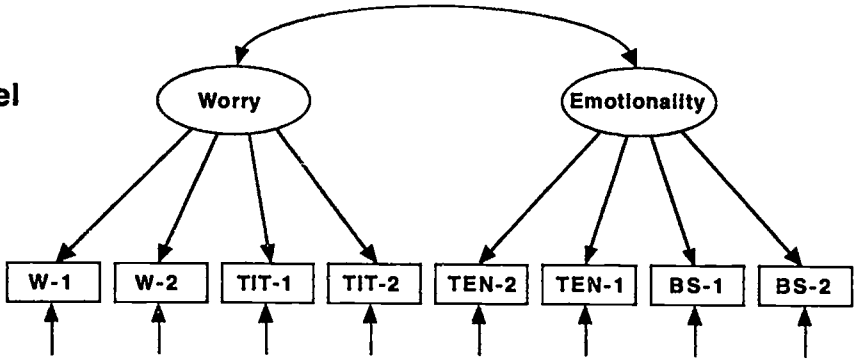
Four Factor Model



Models Based on Spielberger's Theory

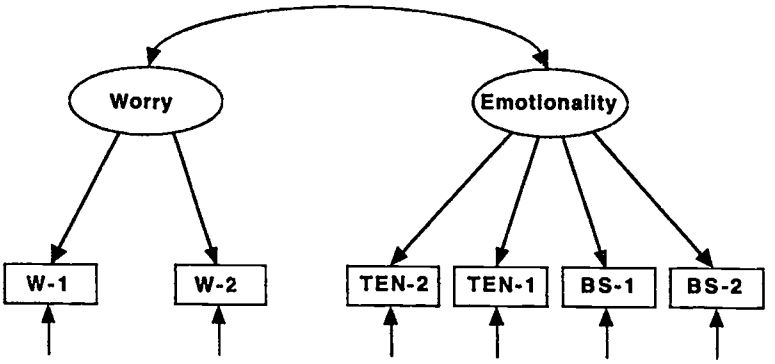
Model 2

Two Factor Model with 8 Indicators



Model 3

Two Factor Model with 6 Indicators



Model 4

Three Factor Model with 8 Indicators

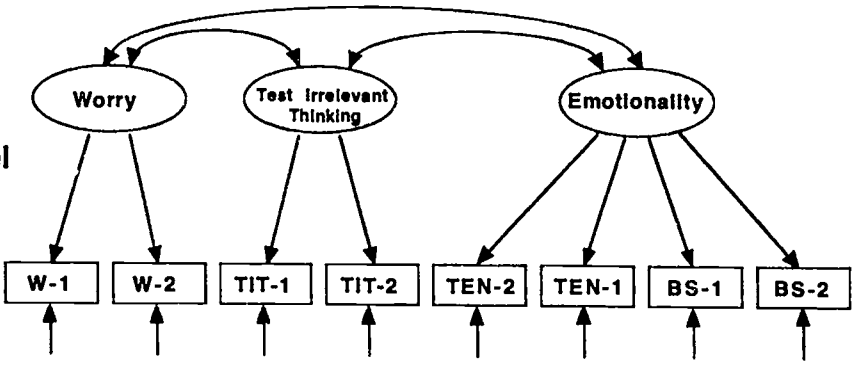


Figure 1. Four Models of Test Anxiety (Specified a priori)