A PSYCHOMETRIC ANALYSIS OF THE NEGATIVE SCALE OF THE ATTRIBUTIONAL STYLE QUESTIONNAIRE WITH AN EYE TOWARDS INVESTIGATING GENDER DIFFERENCES: A COMPARISON OF METHODS

By

CARMEL LAURA KING
B.A., Simon Fraser University, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF EDUCATIONAL AND COUNSELLING PSYCHOLOGY AND SPECIAL EDUCATION

MEASUREMENT EVALUATION AND RESEARCH METHODOLOGY

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2002

© Carmel Laura King, 2002
In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational & Counselling Psychology & Special Education, Measurement Evaluation & Research Methodology

The University of British Columbia
Vancouver, Canada

Date Aug 24, 2002
The present study analyzed the negative scale of the Attributional Style Questionnaire (ASQ), a popular measure of attributional style and correlate of depression (Peterson et al., 1982), using non-parametric item response modeling. This modeling technique was used to determine if items function differentially for males and females, through differential item functioning (DIF) analysis. Past studies of the ASQ have not used non-parametric item response theory (IRT) modeling in their analysis, nor have they determined whether or not items function differentially for males and females. On a methodological note, non-parametric IRT results were therefore compared with results obtained using classical test theory (CTT) statistics. Results indicated that the negative items of each of the three subscales of the ASQ, locus, stability, and globality were found to function well and discriminate best at the extremes. Further, no gender DIF was detected using either non-parametric IRT or CTT methods. Furthermore, conditional reliability was found to be more informative and powerful than a single reliability coefficient. Findings add to the existing literature supporting the theoretical model of the ASQ by utilizing previously unexplored statistical methodologies.
# TABLE OF CONTENTS

Abstract ........................................................................................................... ii
Table of Contents ............................................................................................. iii
List of Tables .................................................................................................... v
List of Figures .................................................................................................. vi
Acknowledgements .......................................................................................... vii

Introduction ................................................................................................... 1
Psychometric Performance Issues ................................................................. 1
Study Purpose ................................................................................................. 2
The Attributional Style Questionnaire ............................................................ 2
Focus of Research on the Attributional Style Questionnaire ......................... 4
Advantages of Item Response Modeling ........................................................ 4
Parametric and Non-Parametric Item Response Models .................................. 5
Study Rationale .............................................................................................. 6

Method .......................................................................................................... 9
Respondents .................................................................................................... 9
Analysis .......................................................................................................... 9

Results .......................................................................................................... 10
Scale Level Analysis- Unidimensionality Analysis .......................................... 10
Item Level Analysis- Non-Parametric Model Fit .............................................. 14
  Differential Item Functioning Analysis ....................................................... 19
  Reliability and Standard Error of Measurement ......................................... 21

Discussion .................................................................................................... 27
List of Tables

Table 1: Results from factor analysis- scale level................................. 12
Table 2: Results from reliability analysis- scale level............................. 13
Table 3: Results from Zumbo (1999b) measure of DIF for ordinal variables as applied to the ASQ negative items................................................. 20
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1:</td>
<td>Non-parametric IRT analysis plots for each of the six items of the negative locus, stability, and globality subscales.</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2:</td>
<td>Reliability and standard error of measurement plots for each of the negative locus, stability, and globality subscales.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 3:</td>
<td>Optimal and observed Standard Error of Measurement (SEM) plots for each of the negative locus, stability, and globality subscales.</td>
<td>24</td>
</tr>
<tr>
<td>Figure C1:</td>
<td>Example of item characteristic curves (ICCs) modeled with a one-parameter logistic model.</td>
<td>46</td>
</tr>
<tr>
<td>Figure C2:</td>
<td>Example of an item characteristic curve (ICC) for a non-parametric item response function.</td>
<td>49</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank Dr. Martin Seligman for the use of his Attributional Style Questionnaire data.
Introduction

As a multi-faceted disorder, depression affects individuals in a number of different ways and in a number of domains (Costello, 1993). As a result, there are numerous depression assessment instruments in use today. Each of these measures differ somewhat in content, response format, and objective; for example, some are self-report measures assessing the severity of depression as a continuum and focus on cognitive, affective, or behavioural symptoms, whereas others only detect the presence or absence of depressive symptoms and are rated by clinicians (Santor & Ramsay, 1998). More specifically, a distinct class of measures has been constructed according to theorists who suggest depression is most beneficially understood by taking into account the causal attributions offered by depressed individuals in explanation for the good and bad events of their lives (Abramson, Seligman, & Teasdale, 1978; Golin, Sweeney, & Shaeffer, 1981; Harvey, 1981; Seligman, Abramson, Semmel, & von Baeyer, 1979). Further, these attributional theorists of depression propose that depressives and non-depressives differ in their causal judgments, and that these differences are closely linked to the existence and range of depressive symptoms (Peterson et al., 1982). One such account of depression is the reformulated learned helplessness model (Abramson et al.) which suggests that the nature of depression is governed by the causal attributions ascribed to aversive events.

Psychometric Performance Issues

Over the years, there has been a considerable increase in the development of measures used in the assessment of depression. However, most of these measures, if not all of them, are subject to a number of enduring questions (Santor & Ramsay, 1998). These questions are mainly concerned with psychometric issues about the performance of these
measures and include: scale discriminability (how effective scales are at detecting differences in depressive severity) and item bias (whether certain groups of individuals endorse items differently). Specifically, these psychometric issues, item bias, and scale discriminability, are best understood through application of item response models. Previously, however, most item analysis techniques used in scale construction and scale evaluation were based on classical test theory (Zumbo, Gelin, & Hubley, 2002).

**Study Purpose**

The purpose of this study is to analyze a popular measure of depressive attributional style, the Attributional Style Questionnaire (ASQ; Peterson et al., 1982) (1) using non-parametric item response modeling, (2) to use this modeling technique to determine if items function differentially for males and females, and (3) to make a comparison of methods at both the scale level and the item level, and between item response theory (IRT) and classical test theory (CTT) statistics.

**The Attributional Style Questionnaire**

The ASQ is a self-report questionnaire designed to measure attributional (or explanatory) style. Attributional style (as an individual differences variable) refers to the habitual ways in which people explain their positive and negative life experiences (Abramson et al., 1978). The ASQ yields scores for individual differences in the tendencies to attribute the causes of good and bad events to internal (versus external), stable (versus unstable), and global (versus specific) factors, as defined by Abramson, Seligman, and Teasdale's (1978) reformulated learned helplessness model. The ASQ is composed of six hypothetical negative events and six hypothetical positive events sampled from the domains of achievement and affiliation (Higgins, Zumbo, & Hay 1999). Respondents are instructed to generate a cause for
each of the 12 events and rate the cause along three 7-point scales representing the locus, stability, and globality causal dimensions. The scale items of the ASQ are anchored so that higher scores indicate more internal, stable, and global attributions (typical of depressed individuals) and lower scores indicate more external, unstable, and specific attributions (typical of non-depressed individuals) (Peterson et al., 1982). Thus, the ASQ generates 36 scores in total, that is, three scores for each of the 12 hypothetical events (see Appendix A for a sample item). Typically, the locus, stability, and globality items are summed (or averaged) across the negative events and positive events separately, to create a locus, stability, and globality score for each type of event (Zumbo, 1999a).

The ASQ was chosen for this study primarily because of its complex scale format. As discussed above, the ASQ is composed of six positive and six negative events each with three questions relating to them, one for each causal dimensions (i.e., locus, stability, and globality). Thus, like many instruments used in psychological research the ASQ is designed with locally (context) dependent item sets. Further, each of the three questions posed for each of the 12 events are rated along a Likert or 7-point scale. These components of the scale format of the ASQ require individual consideration in any psychometric analysis of the ASQ. Thus, the purpose of this study is to examine the ASQ, while taking into consideration its complex format, using non-parametric IRT and to compare this modeling technique to those of CTT statistics. Furthermore, whether test fairness of the ASQ is affected specifically by the gender of the participant(s) is also a consideration. Test fairness, or item bias, will thus also be examined using two differential item functioning (DIF) methods, one from each of IRT and CTT statistics. Past research has not found gender differences for the ASQ, but it is felt that it is always important to look at DIF irrespective of whether gender differences have
been found from previous data. It should also be noted, that both DIF methods to be employed here are relatively new but useful as they will allow us to see their performance with a real data set.

**Focus of Research on the Attributional Style Questionnaire**

Most psychometric research on the ASQ has focused on (a) the predictive utility of the construct (e.g., Peterson, 1991), (b) causal attributions and depression (e.g., Golin et al., 1981; Harvey, 1981; Seligman et al., 1979), (c) developing better measures (e.g., Kinderman & Bentall, 1996; Peterson & Villanova, 1988; Xenikou, Furnham, & McCarrey, 1997), and (d) construct validity (e.g., Higgins et al., 1999). From the social psychology realm, research on attributional style has focused on (a) the effects of situational information (e.g., Mikulincer, 1990), (b) cross-cultural differences (e.g., Nurmi, 1992), (c) self-concept and social functioning (e.g., Bell & McCallum, 1995), and (d) the effects of gender (e.g., Bunce & Peterson, 1997; Greenberger & McLaughlin, 1998; Smith, Hall, & Woolcock-Henry, 2000) (see Appendix B for a review of the literature).

**Advantages of Item Response Modeling**

As of yet, item response models have not been applied to analyses of the ASQ. Item response modeling, however, has many advantages over classical measurement models. More specifically, there are key differences between IRT and CTT that make item response models particularly applicable to the specific psychometric issues affecting depression measures (i.e., item bias and scale discriminability). In particular, the mathematical models employed in item response models acknowledge that the probability of an individual choosing a specific item or option is an inherently stochastic process (Santor & Ramsay, 1998). Briefly, IRT rests on two assumptions: (1) that examinee performance on a test item
can be predicted by a set of factors called traits, latent traits, or abilities, and (2) that the examinee item performance relationship can be described by a monotonically increasing function, or item characteristic curve (ICC) (Hambleton, Swaminathan, & Rogers, 1991), that predicts the item response from the latent variable(s) and item parameter values. Further, the mathematical models employed in IRT assume unidimensionality and local independence. Unidimensionality specifies that only one ability is measured by a set of items in a test. Local independence specifies that, holding constant the abilities influencing test performance, an examinee’s response to any pair of items are statistically independent. Techniques in item response modeling enable researchers to model item responses as a function of a continuum of variation, and thus provide information over the range of the latent variable (Zumbo et al., 2002). The advantage of item response modeling then, when compared to classical methods, is the ability to obtain more information about how the item performs in relation to the scale score (aggregate) (Zumbo et al.). Specifically in relation to item bias and scale discriminability for measures of depressive attributional style (DAS), IRT’s stochastic process acknowledges that effectiveness depends on items being indicators of DAS, that items function differently at different levels of DAS, and that endorsement of any item depends on the individual’s unobserved DAS level.

**Parametric and Non-Parametric Item Response Models**

There are two main forms of item response modeling: parametric and non-parametric (see Appendix C). Parametric item response models conceive the item response function in parametric form (Zumbo et al., 2002); that is, the models contain one or more parameters describing the item and one or more parameters describing the examinee (Hambleton et al., 1991). Non-parametric item response models, however, allow the item response to latent
variable relationship to take on any shape, and thus escape the restrictions of parametric item response models. Of these two main forms of item response modeling, a non-parametric response model was chosen for the current analysis of the ASQ. Non-parametric response models are more appropriately applied to small data sets (e.g., less than 300 respondents) and/or to measures with small numbers of items (e.g., less than 20 items). Therefore, although the ASQ generates 36 scores in total (i.e., three items for each of the 12 hypothetical events), in a scale-to-scale analysis there are only six items per scale (i.e., six negative and six positive events), making it an appropriate measure for a non-parametric model analysis. Also, a non-parametric item response model is beneficial because it will allow exploration of the relationship between the item score and the latent continuum without imposing a strict parametric function; this allows us to investigate whether the item responses increase strictly monotonically with increases in the latent variable (e.g., if the latent variable score associated with a five response is a higher score on the latent variable than a four response).

Rationale for the Study

As a research instrument, the ASQ has been employed predominately in studies of depression (Tennen, & Herzberger, 1985). According to Abramson et al.'s (1978) reformulated learned helplessness model, people who are predisposed to make stable, global, and internal attributions about the causes of bad outcomes are also prone to depression. However, research has not provided similar evidence for attributions about the causes of good events. Internal, stable, and global attributions for good events are weakly correlated with the absence of depression (Peterson, 1991). In addition, it has been shown that attributional style for negative events is a more reliable and valid measure of attributional
performance than attributional style for positive events (Xenikou et al., 1997). Therefore, this non-parametric analysis will include the negative events only.

Possibly one of the most reported issues in the testing literature is the effect of gender on test outcome (as it is related to test fairness) (Hambleton et al., 1991). For many measures, test fairness is affected specifically by the gender of the participant(s). However, this has not been found to be the case for the ASQ. Specifically, past studies on the effects of gender on test outcome on the ASQ have not found gender differences in explanatory style (Bunce & Peterson, 1997; Greenberger & McLaughlin, 1998; Smith et al., 2000). However, these findings have not been substantiated through an IRT analysis. Further, in comparison to other methods used to investigate item bias, IRT provides a unified framework for conceptualizing and investigating bias at the item level (Hambleton et al.). Item bias occurs when examinees from one group (e.g., males) are less likely to endorse an item than examinees of another group (e.g., females) because of some characteristic of the test item that is not relevant to the test purpose (Zumbo, 1999b). Commonly, the term DIF is used to describe the empirical evidence obtained in investigations of bias (Hambleton et al.). DIF occurs when examinees from different groups (e.g., males and females) show differing probabilities of endorsing an item after being matched on the underlying ability that the item is intended to measure (Zumbo, 1999b). It is this matching on the latent variable that makes DIF a much better approach to investigating gender bias than simply comparing item or scale means across the genders. In the current analysis, an approach to non-parametric item response modeling (the TestGraf program) developed by Ramsay (1991, 2000) (see Appendix C) will be used to assess the presence or absence of gender DIF for each item of the domain scales (i.e., locus, stability, and globality). Moreover, as each item of the ASQ is scored according to an ordinal
7-point scale (i.e., Likert-scale) and most of the standard DIF methods are for binary scored items, a new measure of DIF for ordinal variables introduced by Zumbo (1999b) will be applied. This method is advantageous as it tests simultaneously for uniform and non-uniform DIF, and compares these with a measure of effect size (R-squared) for each test. As described by Zumbo (1999b), this DIF method has a natural hierarchy of entering variables into the model, in which the conditioning variable (i.e., total score) is entered first (Step #1), the grouping variable (i.e., gender) is entered second (Step #2), and finally the interaction term (total*gender) is entered last (Step #3). Each of these steps provides a Chi-squared statistic and R-squared value, which are used in the statistical test of DIF and resultant measure of effect size. The DIF computation is the subtraction of the Chi-squared value for Step #1 from the Chi-squared value for Step #3. Similarly, the magnitude of DIF (R-squared value) is computed by subtracting the R-squared value at Step #1 from the R-squared value at Step #3. Subsequently gender DIF will be explored in the negative ASQ items in this manner.

More generally speaking, all IRT models (parametric and non-parametric) include the assumptions of unidimensionality and local independence. Local independence means that once the latent variable is taken into account there is no extra covariation among the items that is not accounted for by the latent variable. However, interpretations based on the ASQ total (negative) score, rather than domain level scores (i.e., locus, stability, and globality), in an IRT analysis, may violate the assumption of local independence. By examining the total score for negative items, it is assumed that one latent variable accounts for the covariation among the items. However, as noted, the questionnaire’s format involves locally dependent item sets, each consisting of the three questions that are posed for each of the six negative
(and six positive) events on the questionnaire (Higgins et al., 1999). These context dependent item sets elicit person covariance and situation covariance; the situation covariance is covariation among the test items over and above that explainable by the latent (person) variable(s) (Higgins et al.). Therefore, in this analysis, only the total scores for each negative dimension will be examined. That is, locus, stability, and globality total scores for each of the negative situations will be computed separately; for these scores, local item independence is not an issue because summation is occurring across situations, not within them. Further, satisfaction of the unidimensionality assumption will be examined for each negative dimension (i.e., locus, stability, and globality) through maximum likelihood one-factor confirmatory factor analysis, and CTT statistics.

The purpose of this paper, then, is to study the psychometric properties (dimensionality, item performance, and gender based DIF) for the negative scale of the ASQ.

**Method**

**Respondents**

The sample is the same as reported in Higgins et al. (1999): 1346 (636 females, 694 males, and 16 who did not indicate their sex) volunteers who completed the ASQ just prior to entering their freshman year at the University of Pennsylvania in 1991. The students ranged in age from 16-24, the mean age was 18.12 years.

**Analysis**

Psychometric analyses are driven by how one computes and reports scores. Scores from the ASQ are computed two different ways: (a) total scores of the positive and negative events; and (b) separate total scores for each of the three dimensions (i.e., locus, stability, and globality) within each of the positive and negative events. The ASQ is comprised of several
situations (six positive and six negative events) each with three items relating to them (one for each of the three dimensions: locus, stability, globality).

For the purposes of this study, the locus, stability, and globality total scores for each of the negative situations will be computed. Analyses will be conducted at both the scale and the item levels. At the scale level, for each of the six negative locus, stability, and globality items, unidimensionality will be examined through maximum likelihood one-factor confirmatory factor analysis, using the PRELIS (version 2.51) program (Jöreskog & Sörbom, 2001) and CTT statistics. At the item level, for the set of the six negative locus, stability, and globality items I will: (a) fit a non-parametric IRT model and compute the item characteristic curves using Ramsay’s (2000) TestGraf program, (b) test for gender DIF using a new measure of DIF analysis for ordinal variables introduced by Zumbo (1999b) and then through TestGraf, and (c) compute conditional reliability and standard error. Finally a comparison of methods will be made at both the scale level and at the item level, and between the IRT and classical methods.

Results

Scale Level Analysis- Unidimensionality Analysis

At the scale level, unidimensionality of each of the three ASQ dimensions was investigated through maximum likelihood one-factor confirmatory factor analysis using the PRELIS (version 2.51) program (Jöreskog & Sörbom, 2001) and CTT statistics. Examination of the root mean square of approximation values resulting from the factor analyses indicated the existence of one factor for each of the negative locus, stability, and globality dimensions (see Table 1). That is, focusing on the locus scale as an example, overall (i.e., irrespective of gender) the factor analysis fit statistic for the one-factor model, RMSEA, indicated a good fit
(i.e., an RMSEA < .05). Furthermore, this good fit held up for the factor analyses of the locus scale for males and females.

Examination of the CTT statistics involved the factor loadings in Table 1 and item statistics in Table 2. Results indicated that, despite the indication of a one-factor model for each of the negative locus, stability, and globality subscales, a few items had poor loadings. For example, locus item 10 had one of the smallest factor loadings overall at 0.34 (only slightly higher than that of stability item 11 at 0.33), and had the lowest item-total correlation (it should be noted, however, that this was only true for female respondents). Further, locus item 19 had a small overall loading and item-total correlation (however, this was only true for male respondents). Item 11 of the stability subscale had the smallest factor loading of any item (poor loading of this item was true for both male and female respondents). Nevertheless, the item-total correlation for item 11 was comparable to that of the other items of the scale. Finally, for the globality subscale, all items had high factor loadings (true for both males and females) and item-total correlations.

Table 2 displays the reliability values calculated using CTT statistics. Overall, and for males and females, reliability was found to be in the range of 0.60 for stability and globality, but in the range of 0.30 for locus. These findings of moderate internal consistency for the stability and globality subscales are consistent with those reported in the literature by several investigators (Golin et al., 1981; Peterson et al., 1982; Seligman et al., 1979). The locus subscale is the least satisfactory and has previously yielded lower reliability scores than the other two dimensions (Tennen & Herzberger, 1985). However, the level of reliability of the negative locus causal dimension may be lower than that of the stability and globality dimensions because of less true scores variance.
Table 1

Factor Analysis - Scale Level

<table>
<thead>
<tr>
<th>ASQ Item Number</th>
<th>Overall</th>
<th>One-Factor Confirmatory Factor Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loading</td>
<td>RMSEA</td>
</tr>
<tr>
<td><strong>Locus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item-4</td>
<td>0.50</td>
<td>0.027</td>
</tr>
<tr>
<td>Item-10</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>Item-13</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Item-19</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Item-22</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Item-31</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td><strong>Stability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item-5</td>
<td>0.48</td>
<td>0.037</td>
</tr>
<tr>
<td>Item-11</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Item-14</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Item-20</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Item-23</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Item-32</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td><strong>Globality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item-6</td>
<td>0.65</td>
<td>0.026</td>
</tr>
<tr>
<td>Item-12</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Item-15</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Item-21</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Item-24</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>Item-33</td>
<td>0.97</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2

**Reliability Analysis - Scale Level**

<table>
<thead>
<tr>
<th></th>
<th>ASQ Item Number</th>
<th>Alpha</th>
<th>Corrected Item-Total Correlation</th>
<th>Alpha Male</th>
<th>Corrected Item-Total Correlation Male</th>
<th>Alpha Female</th>
<th>Corrected Item-Total Correlation Female</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Locus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 4</td>
<td></td>
<td>.3418</td>
<td>.1652</td>
<td>.3262</td>
<td>.1401</td>
<td>.3512</td>
<td>.1936</td>
</tr>
<tr>
<td>Item 10</td>
<td></td>
<td>.0829</td>
<td>.1200</td>
<td>.0366</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 13</td>
<td></td>
<td>.1799</td>
<td>.1729</td>
<td>.1798</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 19</td>
<td></td>
<td>.1470</td>
<td>.1242</td>
<td>.1742</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 22</td>
<td></td>
<td>.1832</td>
<td>.1520</td>
<td>.2102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 5</td>
<td></td>
<td>.6140</td>
<td>.2591</td>
<td>.5954</td>
<td>.2411</td>
<td>.6389</td>
<td>.2922</td>
</tr>
<tr>
<td>Item 11</td>
<td></td>
<td>.2664</td>
<td>.2603</td>
<td>.2745</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 14</td>
<td></td>
<td>.3918</td>
<td>.3762</td>
<td>.4146</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 20</td>
<td></td>
<td>.4049</td>
<td>.4097</td>
<td>.3950</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 23</td>
<td></td>
<td>.3882</td>
<td>.3420</td>
<td>.4446</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 32</td>
<td></td>
<td>.3834</td>
<td>.3701</td>
<td>.4063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Globality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td></td>
<td>.5777</td>
<td>.2307</td>
<td>.5378</td>
<td>.2469</td>
<td>.6074</td>
<td>.2010</td>
</tr>
<tr>
<td>Item 12</td>
<td></td>
<td>.3031</td>
<td>.2263</td>
<td>.3630</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 15</td>
<td></td>
<td>.3395</td>
<td>.3335</td>
<td>.3438</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 21</td>
<td></td>
<td>.3641</td>
<td>.3066</td>
<td>.4219</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 24</td>
<td></td>
<td>.2964</td>
<td>.2493</td>
<td>.3493</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 33</td>
<td></td>
<td>.3727</td>
<td>.3511</td>
<td>.3868</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Number of cases: 1346*
Item Analysis- Non-Parametric Model Fit

At the item level, a non-parametric IRT model was fit to the data and item characteristic curves were computed separately for each of the six negative items of the locus, stability, and globality dimensions. Examination of each item’s response function indicated that the observed ICCs for the six items within each dimension start at the bottom left and increase steadily to the top right corner of each plot (see Figure 1). The slope of those lines at each interval (vertical dashed lines running from left to right across the plot indicate the 5th, 25th, 50th, 75th, and 95th percentiles of the normal distribution, respectively) of the expected score indicates how well items are discriminating among respondents. An item that is functioning appropriately would begin at the bottom left and increase to the top right. The steeper the incline the more discriminating an item; note, however, that the degree of discrimination should reflect where along the X-axis (i.e., negative attributional style) the scale is discriminating best.

For the negative locus, stability, and globality dimensions, the slope of the ICCs illustrates that all items are discriminating well along the continuum of variation of negative attributional style. This is further illustrated by the 95% point-wise confidence limits at various points on the continuum of variation represented by the small vertical solid lines running along each item’s ICC that are grouped more heavily at the corners of each plot. This clustering demonstrates that individuals are most frequently endorsing the extremes of the provided Likert-scale choices (i.e., 1, 2, and 6, 7) rather than the middle points for each dimension (i.e., 3, 4, or 5), indicating the absoluteness of responses. Nonetheless, for locus and globality, endorsement included the entire range from 1-7 whereas for stability,
endorsement ranged from 2-7 for all six items (suggesting that individuals were not willing to say that the cause of the negative event would never again or always be present).
Figure 1

Non-parametric IRT analysis plots for each of the six items of the negative locus, stability, and globality subscales

locus data
stabil data

Item 1

Item Score

Score

Item 2

Item Score

Score

Item 3

Item Score

Score

Item 4

Item Score

Score

Item 5

Item Score

Score

Item 6

Item Score

Score
global data
Differential Item Functioning Analysis

Focusing still at the item level, gender DIF for the domain total score of each negative item of locus, stability, and globality was examined first using a measure of DIF for ordinal variables introduced by Zumbo (1999b) (see Table 3), and then by Ramsay’s TestGraf (2000) index of DIF. As outlined by Zumbo (1999b), Chi-squared and R-squared values were calculated at Step 1, 2, and 3 of the ordinal logistic regression model. The p-values for the resultant two-degree-of freedom Chi-squared test (Step 3 minus Step 1) were also calculated (Chi-squared probability tables are found in most statistical textbooks). In determining whether an item displayed gender DIF, both the two-degree-of-freedom Chi-squared test and the corresponding measure of effect size (R-squared) were considered. In this analysis, in order to classify an item as having DIF, the two-degree-of-freedom Chi-squared test had to have a p-value less than or equal to 0.01 (set at this level because of the number of hypotheses tested) and, in accordance with Zumbo, the R-squared effect size measure had to be at least 0.130. As can be seen through examination of Table 3, obtained p-values and R-squared values did not meet these specified criteria. Therefore, it was determined that there was no significant gender DIF for the negative items of the locus, stability, and globality dimensions. Similarly, DIF statistics obtained through TestGraf did not indicate the existence of gender DIF for any of the negative items of the locus, stability, and globality dimensions. Ramsay’s index of DIF computes a 95% confidence interval for each option (7-point scale) for each of the six items of each dimension. According to this method, if the interval contains zero there is no DIF (Witarsa, in progress). Each option of every item of the locus, stability, and globality dimensions contained zero, and thus, no items were found to display gender DIF.
### Table 3

**Zumbo (1999b) Measure of DIF for Ordinal Variables as Applied to the ASQ Negative Items**

<table>
<thead>
<tr>
<th>ASQ Item Number</th>
<th><strong>Step #1</strong></th>
<th><strong>Step #2</strong></th>
<th><strong>Step #3</strong></th>
<th>DIF COMPUTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Score in the Model</td>
<td>Total Score, and Uniform DIF Variable in the Model</td>
<td>Total Score, Uniform, and Non-uniform DIF Variables in the Model</td>
<td><strong>Step #3 - Step #1</strong></td>
</tr>
<tr>
<td></td>
<td>$\chi^2$ with 1 df</td>
<td>R$^2$</td>
<td>$\chi^2$ with 2 df</td>
<td>R$^2$</td>
</tr>
<tr>
<td><strong>Locus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item-4</td>
<td>453.08</td>
<td>0.28</td>
<td>453.34</td>
<td>0.28</td>
</tr>
<tr>
<td>Item-10</td>
<td>301.88</td>
<td>0.20</td>
<td>301.93</td>
<td>0.20</td>
</tr>
<tr>
<td>Item-13</td>
<td>399.83</td>
<td>0.26</td>
<td>402.01</td>
<td>0.27</td>
</tr>
<tr>
<td>Item-19</td>
<td>373.62</td>
<td>0.24</td>
<td>374.23</td>
<td>0.24</td>
</tr>
<tr>
<td>Item-22</td>
<td>415.66</td>
<td>0.27</td>
<td>415.77</td>
<td>0.27</td>
</tr>
<tr>
<td>Item-31</td>
<td>291.51</td>
<td>0.23</td>
<td>292.14</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Stability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item-5</td>
<td>459.97</td>
<td>0.30</td>
<td>462.18</td>
<td>0.30</td>
</tr>
<tr>
<td>Item-11</td>
<td>438.28</td>
<td>0.29</td>
<td>440.70</td>
<td>0.29</td>
</tr>
<tr>
<td>Item-14</td>
<td>678.62</td>
<td>0.42</td>
<td>679.11</td>
<td>0.42</td>
</tr>
<tr>
<td>Item-20</td>
<td>683.18</td>
<td>0.42</td>
<td>684.12</td>
<td>0.42</td>
</tr>
<tr>
<td>Item-23</td>
<td>682.49</td>
<td>0.41</td>
<td>683.34</td>
<td>0.41</td>
</tr>
<tr>
<td>Item-32</td>
<td>560.74</td>
<td>0.38</td>
<td>561.09</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>Globality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item-6</td>
<td>451.54</td>
<td>0.28</td>
<td>452.06</td>
<td>0.28</td>
</tr>
<tr>
<td>Item-12</td>
<td>525.23</td>
<td>0.32</td>
<td>527.55</td>
<td>0.32</td>
</tr>
<tr>
<td>Item-15</td>
<td>581.08</td>
<td>0.35</td>
<td>581.12</td>
<td>0.35</td>
</tr>
<tr>
<td>Item-21</td>
<td>620.25</td>
<td>0.38</td>
<td>620.42</td>
<td>0.38</td>
</tr>
<tr>
<td>Item-24</td>
<td>468.97</td>
<td>0.30</td>
<td>472.41</td>
<td>0.30</td>
</tr>
<tr>
<td>Item-33</td>
<td>612.75</td>
<td>0.36</td>
<td>614.20</td>
<td>0.36</td>
</tr>
</tbody>
</table>

**Note.** Given the number of hypothesis tests in this table, each DIF test is conducted at a Type I error rate of 0.01.
Reliability and Standard Error of Measurement

In the framework of IRT, reliability and standard error of measurement serve the same role as in CTT statistics. However, as can be seen through examination of the reliability and standard error plots provided in Figure 2, the value of reliability and the standard error in a non-parametric IRT analysis varies along the continuum of the latent trait. For the negative locus, stability, and globality dimensions reliability was highest at the extreme ends of the distribution (i.e., endorsement of options 1, 2, and 6, 7 of the Likert-scale), and lowest in the middle of the distribution (i.e., endorsement of options 3, 4, or 5). Across all three negative dimensions, reliability ranged from 0.52 to 0.82. The standard error was generally lowest at very low and very high total scores (i.e., endorsement of options 1, 2, and 6, 7) and very high for total scores in the middle of the distribution between the possible 6 and 42 total points (i.e., endorsement of options 3, 4, or 5). Across all three negative dimensions, the standard error of measurement ranged from approximately 2.5 to 4. Thus, overall test information and discrimination was highest for the negative scale items of the ASQ at the extreme ends of the distribution.
Figure 2

Reliability and Standard Error of Measurement Plots for Each of the Negative Locus, Stability, and Globality Subscales
The standard error of measurement plots provided in Figure 2, is the standard deviation of TestGraf's best estimate of a subject's trait level taken across the data provided (Ramsay, 2000). These optimal estimates of the standard error of the test score are typically more precise than the score estimated using CTT statistics. To see how the optimal and observed CTT score compare, TestGraf provides a plot containing both the optimal estimate and the observed estimate. As can be seen from the plots provided in Figure 3, the classical (observed) standard error is akin to an average across the continuum of variation whereas the conditional (optimal) standard error varies along the continuum of variation of the latent trait. Thus, more information is provided about the value of standard error at various levels of the latent trait by the optimal curve than the observed curve.
Figure 3

Optimal and Observed Standard Error of Measurement (SEM) Plots for Each of the Negative Locus, Stability, and Globality Subscale Totals

Locus

![Graph showing optimal and observed SEM plots for Locus]
Stability

Standard Error

Observed Score SEM curve

Optimal Score SEM curve

Score
Globality

Standard Error

Optimal score
SEM curve

Observed score
SEM curve

Score

5% 25% 50% 75% 95%
Discussion

The purpose of this study was to analyze the negative scale of a popular measure of depressive attributional style, the Attributional Style Questionnaire (Peterson et al., 1982), (1) using non-parametric item response modeling, (2) to use this modeling technique to determine if items function differentially for males and females, and (3) to make a comparison of methods at both the scale level and the item level, and between IRT and CTT statistics. However, because all IRT models (parametric and non-parametric) include the assumptions of unidimensionality and local independence, initial steps were taken to ensure that these two requirements were met first. Scores from the ASQ are computed two different ways: (a) total scores of the positive and negative events, and (b) separate total scores for each of the three dimensions (i.e., locus, stability, and globality) within each of the positive and negative events. Interpretations based on the ASQ total negative score, rather than domain-level scores (i.e., locus, stability, and globality), in a non-parametric IRT analysis, however, violate the assumption of local independence as the questionnaire’s format involves locally dependent item sets (Higgins et al., 1999). Therefore, only the total scores for each negative dimension were examined to satisfy the assumption of local independence. The unidimensionality assumption was satisfied for each negative dimension (i.e., locus, stability, and globality) through maximum likelihood one-factor confirmatory factor analysis and CTT statistics. Each of these methods clearly indicated the existence of one factor for each of the negative locus, stability, and globality dimensions. Subsequent to satisfying the assumptions of unidimensionality and local independence at the scale level, a non-parametric IRT model was fit to the data at the item level. Results obtained through TestGraf (2000) indicated that the items of all three dimensions of the negative scale of the ASQ function well. Further,
results substantiated the belief that the negative scale discriminates between low and high levels of negative attributional style best at the extremes. Examination of the plots containing the item characteristic curves provided in Figure 1 clearly show groupings at the extreme ends of the curves.

Following both the scale level analysis and the non-parametric IRT item level analysis, gender DIF was analyzed using an ordinal logistic regression method introduced by Zumbo (1999b) and through Ramsay’s (2000) TestGraf program. No items were found to display gender DIF by either method. These findings substantiate the results obtained from the maximum likelihood one-factor confirmatory factor analysis and CTT statistics used initially to establish unidimensionality. DIF detection can be considered the exploration of whether multidimensionality occurs. In this case, the impact of gender as an alternate explanatory dimension was explored and found not to be significant. Therefore, convergent evidence from each method used demonstrates unidimensionality of the individual negative subscales (i.e., locus, stability, and globality) of the ASQ.

The advantages of using non-parametric IRT to calculate conditional reliability over using CTT statistics to calculate a single reliability coefficient are demonstrated. Reliability estimates using CTT tend to be very conservative and do not vary as a function of the latent trait score. The CTT coefficients obtained were lower than their associated conditional reliabilities on each of the three subscales. Thus, conditional reliability is a more powerful way of measuring test quality.

In summary, for the sample used here, the negative locus, stability, and globality subscales of the ASQ were found to be unidimensional using both classical test theory and non-parametric item response theory statistics. Further, no gender DIF was detected using
either non-parametric IRT or the Zumbo (1999b) ordinal logistic regression method. Furthermore, DIF has proved to be a much better approach to investigating gender bias than simply comparing item or scale means across genders. The difference between DIF through IRT and CTT is that IRT DIF methods match the groups on the underlying latent variable, and thus consider the ICCs for the two groups on the same item along the continuum of variation. Finally, conditional reliability values obtained using TestGraf were found to be more informative and powerful than a single reliability coefficient obtained through CTT statistics. Thus, these findings add to the existing literature supporting the theoretical model of the ASQ by utilizing previously unexplored statistical methodologies.

Implications

The findings of this study have several implications for other researchers. Firstly, findings of this study have substantiated past findings on the psychometric properties of the ASQ; i.e., that the negative scale of the ASQ is unidimensional and reliable within its three dimensions and not subject to gender DIF. Secondly, that the application of a non-parametric IRT model in the analysis of a scale with a complex format has many benefits, including the ability to address the issues of item bias and scale discriminability at the latent trait level. Finally, that researchers interested in studying negative attributional style or other correlates of depression can confidently make use of the ASQ in various populations.
References


Appendix A

Attributional Style Questionnaire: Example Item

The following example illustrates the nature of the questions for one situation (one question is provided for Locus, Stability, and Globality respectively):

You have been looking for a job unsuccessfully for some time.

Write down one major cause ________________________________

1. Is the cause of your unsuccessful job search due to something about you or to something about other people or circumstances? (circle one number)
   Totally due to
   other people  1  2  3  4  5  6  7  to me
   or circumstances

2. In the future when looking for a job, will this cause again be present? (circle one number)
   Will never again
   be present  1  2  3  4  5  6  7  Will always
   be present

3. Is the cause something that just influences looking for a job or does it also influence other areas of your life? (circle one number)
   Influences just
   this particular  1  2  3  4  5  6  7  all situations
   situation       in my life
Appendix B

Literature Review

Attributional Style and the ASQ

Attributional style is commonly referred to in the literature according to Abramson, Seligman, and Teasdale's (1978) reformulated learned helplessness model (e.g., Golin, Sweeney, & Shaeffer, 1981; Harvey, 1981; Higgins, Zumbo, & Hay, 1999; Kinderman & Bentall, 1996; Mikulincer, 1990; Peterson et al., 1982; Peterson & Villanova, 1988; Seligman, Abramson, Semmel, & von Baeyer, 1979; Smith, Hall, Woolcock-Henry, 2000). In fact, it is fair to say that Abramson et al.'s reformulated learned helplessness model has created a great deal of interest in the role of attributional style in depression. Currently, the ASQ is the most commonly used instrument in the assessment of attributional style (Kinderman & Bentall). Despite its frequent use and application, however, the ASQ has been reported to have only satisfactory reliability (Peterson et al., 1982). Further, studies have also noted that the ASQ has low reliability within its individual dimensions (Peterson & Villanova), especially the internality (i.e., locus) dimension (Kinderman & Bentall). In a specific attempt to examine the low internal reliability of the internality dimension, Kinderman and Bentall developed the Internal, Personal, and Situational Attributions Questionnaire (IPSAQ). According to Kinderman and Bentall, three distinct attributional loci could be identified within the ASQ's internality dimension. These three loci taken from the ASQ's internality dimension (i.e., internal, personal, situational) make up the attributional categorizations on the IPSAQ. Results of Kinderman and Bentall's study revealed acceptable levels of internal reliabilities that were substantially superior to those reported for the internality subscale of the ASQ.
Further attempts to improve the reliability of the ASQ have been made by Peterson and Villanova (1988) and Xenikou, Furnham, and McCarrey (1997). Kinderman and Bentall added more negative items (explanatory style for bad events is often examined independent of explanatory style for good events (Peterson, 1991)) to the ASQ in an attempt to increase its internal consistency. Their Expanded Attributional Style Questionnaire (EASQ) had internal reliabilities in the individual dimensions that were substantially higher than those for the ASQ. More specifically, the internal consistency of each dimension of the ASQ ranges from .4 to .7, which was increased to .7 to .9 respectively (for negative events only) in Peterson and Villanova's expanded version. However, the result was hardly surprising given the fact that ability to assess "style" increases with the number of events included in an assessment of it (Peterson). Further, increases in reliability can occur in any scale, containing items positively correlated to one another, with the addition of more items (Peterson). Thus, as Peterson and Villanova noted, by increasing the number of negative items from 6 to 24 the internal consistency of the ASQ was improved, but the result was a psychometric certainty, and therefore only added support to the notion that the ASQ should be improved. Similarly, Xenikou et al. proposed a more reliable measure of attributional style by focusing on the negative events for the dimensions of stability and globality. Results of their study indicated that negative outcomes load on two correlated factors (i.e., stability, and globality) not three, as initially proposed by Abramson et al. (1978), that the internality dimension was not consistent across situations, and that both negative stability and negative globality were reliable measures.

In addition to examination of the internality dimension of the ASQ, the stability of attributions has also been of interest. For instance, Mikulincer (1990) assessed the role of
situational factor involvement in forming causal attributions, and examined the proposition made by Alloy and Tabachnick (1984) that there is a joint influence of situational information and attributional style. Using a Hebrew version of the ASQ, Mikulincer divided students according to their attributional style for negative events into stable, undefined, and unstable attributors. Mikulincer found that all participants made more stable attributions when situational cues indicated a stable cause than when an unstable one was indicated. Also, attributional style was found to influence future expectancies and quality of performance following failure. Conversely, Nurmi (1992) examined cultural differences in causal attributions and self-serving bias as they pertain to all three of Abramson et al.'s (1978) causal dimensions. In a comparison of American and Finnish students, Nurmi showed that Americans used self-serving attributional bias to a greater extent than the Finns. That is, in accordance with Abramson’s et al.’s reformulated learned helplessness model, Nurmi found that Americans attributed good outcomes to more internal, stable and global factors than the Finns, with the reverse being true for bad outcomes (Nurmi).

The relationship between attributional style and self-concept and social functioning has also been of interest. In a study by Bell and McCallum (1995) a measure of children’s attributions, the Student Social Attribution Scale (SSAS), was developed. The authors proposed a six-factor structure for the SSAS, originating from the combination of two facets in the school social domain: outcome (success; failure); and attribution (ability; effort; external causes), to measure students’ perceptions of causes of their school-related social success and failure (Bell & McCallum). Factor analysis provided support for the six hypothesized factors; however the external scales were psychometrically weaker than the ability and effort scales. Further, results indicated that attributions appeared to be related to
self-concept and social skills in the school setting (Bell & McCallum). “More specifically, data supported hypothesized relationships among students’ self-attributions for social success and failure” (Bell & McCallum, p.271). That is, ability and effort attributions for social success were positively related to self-concept and social functioning, with the reverse being true for ability and effort attributions for social failure (Bell & McCallum).

Support for the Reformulated Learned Helplessness Model

Related to studies that have investigated the individual dimensions of attributional style proposed by Abramson et al. (1978), are studies which have sought to add support for the reformulated learned helplessness model or make predictions in conjunction with it. For example, Harvey (1981) predicted that, in addition to Abramson et al.’s hypothesis that internal, stable, and global attributions (i.e., negative attributional style) lead to depression, depressed individuals would also attribute negative events to more controllable causes, and that controllable causal attributions would be more closely related to independent judgments of the controllability of events. Results offered support for the learned helplessness model; however, the further predictions made by Harvey were only partially supported. Similarly, Smith et al. (2000) added support for Abramson et al.’s reformulated learned helplessness model in their study of optimistic and pessimistic explanatory style. Examination of dimension composite scores on the ASQ enabled the authors to conclude that optimistic explanatory style is evidenced by more external, unstable, and specific attributions, whereas pessimistic attributional style is evidenced by more internal, stable, and global attributions (Smith et al.). Further, Seligman et al. (1979) found that relative to non-depressed students, depressed students attributed bad outcomes to more internal, stable, and global factors. Similarly, Higgins et al.’s (1999) factor model demonstrated that correlations between the
attributional style factors are consistent with the idea of depressive attributional style. Conversely, Golin et al. (1981) showed that internal attributions do not play a causal role in depression, and that only stable and global attributions might act as causes.

In general, studies on attributional style have added support for Abramson et al.'s (1978) model, and thus have shown that attributions for bad outcomes do play a causal role in depression. However, there seems to be general consensus throughout the literature that the internality dimension is the least reliable, which points to a need for revision of that dimension.

Attributional Style and Depression

A great deal of past research on Abramson et al.'s (1978) reformulated learned helplessness model, and attributional style in general, has included a comparison of attributional style and depression as they are measured by specific assessment instruments. For example, Kinderman and Bentall (1996) noted that, as predicted by the attributional accounts of depression, depressed patients have been known to make excessively internal, stable, and global attributions for the negative events of the ASQ. However, the exact nature of the links between attributional style and psychopathology has not been determined (Kinderman & Bentall). Therefore, the authors made comparisons between the Beck Depression Inventory (BDI) and attributional style as measured by their IPSAQ. Their results indicated that internal attributions (authors focused only on the internality dimension) for negative events were more closely associated with low mood than were personal or situational external factors (Kinderman & Bentall). In a similar study, Harvey (1981) examined the relationship between attributional style (as measured by a Life Stages Questionnaire modeled after the ASQ), controllability, and judgments of controllability of
events, and depression. Results indicated that predicted differences between depressed and non-depressed students (as measured by the BDI) were found for the internal attributional dimension, and that depressed students attributed negative events to more controllable causes than non-depressed students.

In one early study of attributional style and depression, Seligman et al. (1979) compared scores on an attribution style scale with those of two measures of depression, the BDI and the Multiple Affective Adjective Checklist (MAACL) (Zuckerman & Lubin, 1965). Correlational data from this study showed that the more depressed the participants were on both the BDI and the MAACL, the greater were their ratings of the internal, stable, and global attributional causes of bad outcomes. Similarly, in a study by Peterson and Villanova (1988) it was shown that all three individual dimensions of attributional style (internal, stable, and global) correlate positively with depressive symptoms as measured by the BDI. However, in an earlier study by Golin et al. (1981), which compared data from the ASQ and BDI, it was determined that predispositions to make stable or global (but not internal) attributions for bad outcomes might be a cause of depressive symptoms.

The reformulated learned helplessness model (Abramson et al., 1978) is an attributional account of depression. Thus, it is no wonder that many studies of attributional style (that have not done a comparison of attributional style and depression directly) make reference of their findings as they pertain to depression. For example, in their study outlining the development of the ASQ, Peterson et al. (1982) noted that an attributional style in which internal, stable, and global attributions are offered for bad events is associated with depressive symptoms. Further, in an investigation of cross-cultural differences in self-serving bias in Americans and Finnish students, Nurmi (1992) pointed out that results of his study
showed some similarity with those reported by Seligman et al. (1979) for mildly depressive American students. Specifically, Nurmi reported that Americans had a higher self-serving attributional bias, whereas the Finnish students tended to make more frequent use of a defensive-pessimistic strategy that is lacking in self-serving bias. Similarly, Smith et al. (2000) examined pessimistic and optimistic attributional style in secondary vocational teachers and noted that, according to Seligman (1990), pessimistic explanatory style is exemplary of individuals who give internal, stable, and global explanations for bad events. Finally, Higgins et al. (1999), in their study of construct validity of attributional style, showed correlations between attributional style factors that were consistent with depressive attributional style. Overall, the literature examined substantiated the existence of a causal link between attributional style and depression.

Attributional Style and Gender

"The ASQ was not designed as a clinical tool" (Tennen & Herzberger, 1985 p.22). Nevertheless, the ASQ has been useful in understanding the role of attributional style in many areas of clinical (and psychological) research. One such area is in research examining gender and sex role differences in causal attributions (Tennen & Herzberger). For example, a study by Greenberger and McLaughlin explored sex differences in attachment, coping, and explanatory style (as measured by the ASQ), and found that males and females did not differ in explanatory style. In a study by Smith et al., it was found that male and female vocational teachers had optimistic explanatory styles, and that both had similar explanatory styles toward the negative and positive events of the ASQ. Similarly, Bunce and Peterson (1997) found no significant differences between males and females on any of the three dimensions of the ASQ, for either the negative or positive events. Overall, past research examining
gender and sex role differences in causal attributions, which have employed the ASQ, have not found sex differences in explanatory style.

Construct Validity and Attributional Style

Research investigations that examine the influence of context on responses to questions have been of great interest in the literature (Steinberg, 2001). Of specific interest is that context can influence the attachment of meaning to an item and thus influence the selection of a response (Tourangeau, 1999). However, few studies of the ASQ have addressed the fact that the format of the questionnaire involves context dependent item sets (Steinberg & Thissen, 1996). Like many instruments used in psychological research the ASQ is designed with a stem followed by several (three) questions presented as a set (Steinberg, 2001). Most researchers whose analyses involve the ASQ create a composite of internality, stability, and globality, for the negative events and/or the positive events, or the whole scale, and treat the composite score as an index of explanatory style (Peterson, 1991). However, the context dependent item sets (CDIS) in the ASQ elicit person covariance and situation covariance. The situation covariance is covariation among the test items over and above that explainable by the latent (person) variable(s) (Higgins et al., 1999). Thus, this extra covariation must be taken into account in any analysis of the ASQ total score. In a study by Higgins, Zumbo, and Hay this extra covariation was described in relation to the format of the ASQ as CDIS and its effect was tested using four models of attributional style, with and without the CDIS included in each model. Results indicated that a three-factor model (i.e., internality, stability, globality) that included CDIS had adequate fit. Further, Higgins et al.’s findings demonstrated “that CDIS are introducing extra covariation in people’s responses to the ASQ” (p.818). Therefore, attributional style is situational. Furthermore, attributions about
hypothetical events (such as those posed on the ASQ) are not related to actual events (Peterson, 1991). Nevertheless, it is not possible to negate the impact of situational characteristics of attributional style. As such, researchers using the ASQ must take into account the extra covariation elicited by person covariance and situation covariance; and as shown by Higgins et al., modeling CDIS allows for this.
Appendix C

Item Response Theory, TestGraf, and Differential Item Functioning

Parametric and Non-Parametric IRT

There are two main forms of item response modeling: parametric and non-parametric. Parametric item response models conceive the item response function in parametric form (Zumbo, Gelin & Hubley, 2002); i.e., the models contain one or more parameters describing the item and one or more parameters describing the examinee (Hambleton, Swaminathan, & Rogers, 1991). For example, the one-parameter Rasch model, or the two or three parameter logistic models are given by the following equations respectively:

\[ P_i(\theta) = \frac{e^{\lambda(\theta-b_i)}}{1 + e^{e-b_i}} \]

Where:
- \( P_i(\theta) \) is the probability that a randomly chosen examinee with ability \( \theta \) answers item i correctly. \( P_i(\theta) \) is and S-shaped curve with values between 0 and 1 over the ability scale.
- \( b_i \) is the item i difficulty parameter.
- \( n \) is the number of items in the test.
- \( e \) is a transcendental number (like \( \pi \)) whose value is 2.718 (correct to three decimals).
- \( D \) is a scaling factor introduced to make the logistic function as close as possible to the normal ogive function. \( D = 1.7 \)
- \( a_i \) is the item discrimination parameter.
- \( c_i \) is the pseudo-chance level parameter; incorporated into the model to take into account performance at the low end of the ability continuum, where guessing is a factor in test performance on selected-response (e.g., multiple choice) test items.

(From Hambleton et al., pp.12-17)

Clearly the one, two, and three parameter models are based on restrictive assumptions; that is, they assume that knowing the degree of discrimination, \( a_i \), difficulty, \( b_i \), or \( c_i \), chance of guessing on items are sufficient to account for the relation between responses made by individuals and the underlying construct \( \theta \) (Santor & Ramsay, 1998). Therefore, the appropriateness of these assumptions depends on the nature of the data, the importance of the
intended application (Hambleton et al.), and the “correctness” of the mathematical model chosen (Santor & Ramsay). Moreover, in order to get a clear understanding of research results, parametric models are best suited for measures with items constructed with a logistic model in mind. Non-parametric item response models, on the other hand, do not contain parameters, and thus, escape the restrictions of parametric item response models.

Item response model techniques allow researchers to model item responses as a function of a continuum of variation (Zumbo et al., 2002). For parametric models, how responses to items change across different levels of any trait, latent trait, ability or condition, such as attributional style, is described by the ICCs (Santor & Ramsay, 1998). In parametric models the item characteristic function is logistic, specifying that as the level of the trait increases, the probability of a correct response to an item increases (Hambleton et al., 1991). Thus, option characteristic curves represent the probability of a specific response option being endorsed or observed at any level of the trait, ability, or condition (Santor & Ramsay). An example of option characteristic curves modeled with a one-parameter logistic model is illustrated in Figure C1.
Figure C1

*Example of Item Characteristic Curves (ICCs) Modeled with a One-Parameter Logistic Model*

One-Parameter ICCs for Four Typical Items:
For non-parametric models, however, the ICCs can correspond to graded statements or to levels of a Likert-scale, such as on ASQ. Therefore, non-parametric item response models estimate response curves directly, and thus, escape the restrictions imposed by parametric models. Moreover, in contrast to parametric models, non-parametric item response models make no a priori assumptions about the underlying distribution of responses or their corresponding relationship to the underlying trait, ability or condition being measured (Santor & Ramsay), and thus, they are extremely flexible.

Non-Parametric IRT and TestGraf

A non-parametric response model was chosen for the current analysis of the ASQ negative scale. Non-parametric response models are more appropriately applied to small data sets (e.g., less than 300 respondents) and/or to measures with small numbers of items (e.g., less than 20). Therefore, although the ASQ generates 36 scores in total (i.e., three items for each of the 12 hypothetical events), in a scale-to-scale analysis there is only six items per scale (i.e., six negative and six positive events), making it an appropriate measure for a non-parametric model analysis. In particular, an approach to non-parametric item response modeling developed by Ramsay (1991, 2000) was applied in analyzing the ASQ negative scale. Ramsay’s non-parametric kernel-smoothing techniques (i.e., Gaussian Kernel) are implemented with the application of the TestGraf program (Ramsay, 2000). In contrast to logistic response models, non-parametric kernel-smoothing techniques estimate $P_{jm}(\theta)$ (the probability of individuals endorsing option $m$ from item $j$ as a function of $\theta$) at a set of equally spaced evaluation points ($q_i$) selected from the distribution of standard normal scores (Santor & Ramsay, 1998). Thus, instead of conceiving of the item response function as a smooth increasing function with parametric form (Zumbo et al., 2002) (e.g., one-parameter
model displayed in Figure C1), the non-parametric function flexibly fits to the data; which is particularly useful when response characteristic curves are expected to change rapidly, for example, with changes in depressive severity (Santor & Ramsay). Figure C2 is an example of an item response function computed through non-parametric item response modeling, using Ramsay's TestGraf computer program (Zumbo et al.). The item characteristic curve for the non-parametric item response function is the solid line depicted in Figure C2. The small vertical solid lines running along the ICC indicate the 95% pointwise confidence limits (these are not 95% confidence limits for the entire curve, but rather confidence limits at particular points on the continuum of variation) (Zumbo et al.). The vertical dashed lines running left to right across the plot indicate the 5th, 25th, 50th, 75th, and 95th percentiles of the normal distribution, respectively.
Figure C2

Example of an Item Characteristic Curve (ICC) for a Non-Parametric Item Response Function
Basically, the objective of TestGraf is to display the relationship between the trait (ability) level of an examinee (θ) and the probability that they choose various options for each item (Ramsay, 2000). The graphical displays produced by TestGraf, for non-parametric response modeling, do not display the numerical values of the observed total score; the goal of the item analysis is to work at the latent variable level. In Figure C2 (and for the purpose of this analysis), the expected scale score was used. In calculating the expected scale score, TestGraf replaces the observed scores by their ranks, and then replaces the ranks by the corresponding standard normal quantiles (z-scores) prior to applying the Gaussian Kernel smoothing for the non-parametric regression. This is permissible because, as Ramsay indicates, one cannot measure a latent variable like ability (θ) in the usual sense. That is, one can only know its values to within any transformation that preserves the rank order among the observed total scores, called a monotone transformation. Although the class of monotone transformations is limited, TestGraf provides several graphical display options for the abscissa (X-axis) (e.g., expected total score, standard normal quantiles, and standard error total score). Therefore, the user has the option of choosing a display variable appropriate to the data set and characteristics of the test being analyzed.

**Differential Item Functioning**

Possibly one of the most reported item bias issues in testing literature is the affect of gender on test outcome (Hambleton et al., 1991). Specifically in relation to the ASQ, past studies have not found gender differences in explanatory style (Bunce & Peterson, 1997; Greenberger & McLaughlin, 1998; Smith et al., 2000). However, these findings have not been substantiated through an IRT analysis; previously, examinations of bias for the ASQ have been based on CTT techniques. Further, in comparison to other methods used to
investigate item bias, IRT provides a unified framework for conceptualizing and investigating bias at the item level (Hambleton et al.). Commonly, the term DIF is used to describe the empirical evidence obtained in investigations of bias (Hambleton et al. 1991). DIF occurs when examinees from different groups (e.g., males and females) show differing probabilities of endorsing an item (Zumbo, 1999b). Basically, the difference between the DIF techniques employed through an IRT analysis and those of a CTT analysis is that DIF matches the groups of interest on the latent variable. Thus, from the IRT framework, one considers the ICCs for the two groups being compared on the same item along the continuum of variation (Zumbo & Hubley, 2002). Generally, in investigations of DIF, the ICCs can differ in two ways. First, the curves can differ in terms of their threshold (i.e., difficulty) parameter (Zumbo & Hubley). "If the parameters of the two item characteristic functions are identical, then the functions will be identical at all points" (Hambleton et al. p.110), if they are not the curves will be displaced by a shift in their location on the theta continuum of variation (uniform DIF) (Zumbo and Hubley). Second, the ICCs may differ on discrimination and hence the curves may be seen to intersect (non-uniform DIF) (Zumbo & Hubley). Alternatively, another approach is to compare the area between the two item characteristic functions (curves) (Hambleton et al.). If the area between the curves is not zero then it can be concluded that DIF is present. However, when DIF is found to have occurred further analyses are then needed to determine whether it is a result of item bias or item impact. Item bias, occurs when DIF exists because of some difference (e.g., characteristic of the test or testing situation) that is irrelevant to the underlying trait being measured (Zumbo & Hubley). Item impact, on the other hand, occurs when DIF exists because of true differences in the underlying trait being measured by an item between the two groups being compared (Zumbo.
& Hubley). Thus, although detection of DIF is useful for the identification of problematic items, alone it does not indicate whether an item has bias or impact; further study is necessary to determine the nature of the DIF being displayed.