Differential Item Functioning on the International Personality Item Pool's Neuroticism Scale

Nadine LeBarron McBride

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Robert J. Harvey, Chair

Martha Ann Bell

T. W. Bonham

Kevin D. Carlson

November 11, 2008

Blacksburg, VA

Keywords: differential item functioning, DFIT, IPIP, Neuroticism

Differential Item Functioning on the

International Personality Item Pool's Neuroticism Scale

Nadine LeBarron McBride

ABSTRACT

As use of the public-domain International Personality Item Pool (IPIP) scales has grown significantly over the past decade (Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006) research on the psychometric properties of the items and scales have become increasingly important. This research study examines the IPIP scale constructed to measure the Five Factor Model (FFM) domain of Neuroticism (as measured by the NEO-PI-R) for occurrences of differential functioning at both the item and test level by gender and three age ranges using the DFIT framework (Raju, van der Linden, & Fleer, 1993) This study found six items that displayed differential item functioning by gender and three items that displayed differential item functioning by age. No differential functioning at the test level was found. Items demonstrating DIF and implications for potential scale revision are discussed.

Acknowledgments

I must extend my deep gratitude and appreciation to my advisor, R.J. Harvey. This dissertation is largely due to his faith in my ability and perseverance. He encouraged me when I needed it and pushed me when I may not have wanted to hear it, but was very much needed.

Sincere thanks to my committee for their willingness to stick with me during what became a long process. I proposed this dissertation with every intent of completing it in a timely manner. Three years, a job change, and a child, later, they were enthusiastic about seeing me through to completion. I could not appreciate that more. Their ideas and discussions helped me tremendously.

I am also appreciative of Bart Craig at North Carolina State University for his assistance with the computer programs used in this study and patience in answering my questions concerning the resulting output.

I would not be where I am today without the support of Gayle Kennedy. Her diligence and willingness to go the extra mile on my behalf is extremely appreciated.

I've been fortunate to work closely with two wonderful psychometricians during my early professional career. Jim Penny and Laura Kramer have challenged me and taught me to apply "book knowledge" to "real world" situations and come up with answers that were not only technically sound but practical. I am grateful for their guidance, support, and willingness to let me bounce ideas off of them throughout this process.

Lastly, my husband, Andy, who has been by my side since I was in high school. He has pushed me to succeed and supported me throughout my entire graduate career. His belief in me and patience and understanding often provided me the confirmation I needed.

Table of Contents

List of Tables

List of Figures

Introduction

For decades mean group differences on test scales have led people to wonder if such a scale measures accurately or if it is biased against the group scoring lower. Although one cannot assume from group mean differences that a scale is biased, given that a decision that may be made from scores on the test scale, whether for employment selection, college admissions, or other purposes for which a scale may be used, the possibility should be explored. Given that these decisions have very real, perhaps substantial, impact on an individual's life, test developers must make every reasonable effort to ensure that a test is not biased against any particular subgroup of individuals who may take the test.

Although debate on potential test bias has been more prevalent amongst constructs such as cognitive ability, recently several research studies have begun to explore potential test bias in non-cognitive measures such as personality (e.g. Forrest, Lewis, & Shevlin, 2000; Huang, Church, & Katigbak, 1997; Reise, Smith, & Furr, 2001; Robie, Zickar, & Schmit, 2001; Smith, 2002). Specifically, gender group differences in numerous measures of the personality trait of neuroticism have lead researchers to examine such measures for potential bias in the items contained in the measures (Reise, Smith, & Furr, 2001; Smith & Reise, 1999). Given that measures of neuroticism have been found to negatively correlate with job satisfaction (Furnhman & Zacherl, 1986), management skills (Furnham & Mitchell, 1991), symptoms of job pressure and dissatisfaction (Kirkaldy, Thome & Thomas, 1989), positively correlate with work-family conflict (Rantanen, Pulkkinen, & Kinnunen, 2005) as well as predict aspects of job burnout (Godderd, Patton, & Creed, 2004; Zellars, Hochwarter, & Perrewé, 2004), it is important to determine if the items or scales used in such research may function differently for different groups of individuals.

This study extends such research examining potential differences in functioning to the emotional stability/neuroticism scale of a new personality measure based on the Five Factor model, the International Personality Item Pool, using the DFIT framework proposed by Raju, van der Linden, & Fleer (1995). It also takes the additional step of examining the measure for bias at the test level, not just item level.

Differential functioning at either the item or test level has clear implications for the International Personality Item Pool (IPIP). First and foremost, the IPIP is a public-domain, international scientific effort, supported by research of many scientists dedicated to the continuous improvement of the measurement of human personality. This research is used for further development and refinement of the IPIP (Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006). The discovery of items displaying DIF permits the investigation into such items to explore why such DIF may occur. Additionally, by employing the unique DFIT framework, one can explore the effect of removal of items displaying DIF to the overall DTF.

Definitions of Test Bias and Differential Functioning

When mean scores of two population subgroups differ on a measure there are two possible general explanations. The first is that the test accurately assesses the level or ability and that the two groups truly differ in level of ability. The other possible explanation is that the level of trait or ability is consistent across the two groups but that the test is not accurately measuring ability for one of the groups. In the second case, the test is not measuring the same construct for each of the two groups and the test may be referred to as biased.

The term "bias" has been defined in multiple ways in the past. Jensen (1980) and Cleary (1968) define test bias in terms of systematic error in the predictive validity that is associated with group membership. Court rulings have also defined bias as differences in passing rates between subgroups (Allen v. Alabama State Board of Education, 1985), confounding measurement bias with potentially true differences in proportion correct. However, measurement bias is more accurately defined as differences in the test scores of groups with identical ability. At the item level, bias is the differences in the probability of answering a question correctly, given identical ability (Stark, Chernyshenko, & Drasgow, 2004). Thus, the item in question is differentially more difficult for one group than the other.

A distinction needs to be made between the use of the words “bias” and “fair.” Although the term fairness is often muddled with the term bias (Shepherd, 1982; Reynolds, 1982; Green, 1975), the use of the term fair includes the use of the test, such as in a selection context, while the term bias should be used to refer to an inherently statistical property that is contained within the test. Thus, issues of fairness, such as the misuse of a test in a selection context leading to discrimination, cannot be determined purely through psychometric research, whereas determinations of statistical bias can.

Attempts at clarifying this semantic confusion have been made through the use of the term differential item functioning (DIF) in the study of test and item bias. Although often used synonymously with the term bias (Camilli & Shepherd, 1994), the use of the term differential item functioning is limited to the observation that two groups, which are matched on level of ability, do not perform identically on a given item (Angoff, 1993; Penfield & Lam, 2000). The groups of interest are typically referred to as the reference group and the focal group. Usually, the reference group is the majority group in the comparison while the focal group is the minority.

Essentially, the two groups either differ on their true score and the measure is accurately reflecting these differences or the two groups have similar true scores and the measure does not accurately reflect their standing on the latent trait. In the first case, the measure is said to have measurement equivalence between the groups and meaningful group differences can be addressed. When the second scenario occurs, the test or scale is not measuring the same latent trait for the two groups. Then, the test is said to be *differentially functioning* between two groups.

The term differential item functioning is intended to describe a statistical observation only, deplete of any social meaning. When one defines bias as implying unfairness to one group, it is important to emphasize that the terms differential functioning and bias are not synonymous. When bias is defined as a statistical property alone, not to be confused with fairness, the two terms can be used interchangeably as they are throughout this paper.

Judgmental Methods of Detecting Item Bias

Upon the realization that mean group differences existed, early efforts to identify biased items began with judgmental methods. In this method, items are reviewed for face and content validity by experts (Tittle, 1982). These judgments are used to correct or eliminate potentially biased items by focusing on eliminating item content considered to be stereotypical of a subgroup or derogatory in nature. They are also used to ensure fair representation of women and minorities, as well as ensuring inclusion of material thought to be of equal familiarity of or experience with the content by various subgroups.

Major test publishers and other developers of examinations have developed guidelines for the review of items for content that may be viewed as biased or offensive (e.g. Macmillan, 1975; McGraw-Hill, 1968). Although statistical methods are now more often considered primary a method of identifying biased items, judgmental methods continue to be routinely incorporated during the item and test development process. Educational Testing Services (ETS), for example, includes trained reviewers in item development as a part of their intensive sensitivity review process (Ramsey, 1993). A comprehensive review of items by experts, both internal and external to the test development process, for content that may be biased provides the additional benefit of increasing face validity and the perception that the test is "fair."

Overview of Statistical Methods for Detecting Bias

There have been two main statistical approaches to detecting bias since bias has become a point of interest in testing: one using an external criterion and one using a criterion internal to the test. Predictive validity models, such as the Clearly model (Cleary, 1968), use an external criterion and examine regression lines to determine if a common regression line for the population underpredicts or overpredicts for members of one group (Petersen & Novick, 1976). If the regressions of the criteria on test scores are compared and if found not to equitably predict performance, it can be hypothesized that the slope of the regression line is different for the two groups, such as positive for the majority group and zero for the minority group, or that the slopes of the regression line are the same but the intercept is higher for the minority group than for the majority group, leading to underprediction of performance for the minority group. Multiple regression analysis can be used to determine the form which occurs. Regardless, it is assumed to be due to an undesirable causal factor other than the criterion, indicating test bias.

Predictive validity models have frequently been used in investigation of "fairness" of tests such as those used in an employment selection setting (Sireci & Geisenger, 1998). However, numerous technical problems have been found with this method as it measures test bias. Not only is the criterion assumed to be unbiased (Petersen & Novick, 1976) but also the model itself may be statistically biased against minorities (Terris, 1997). In their review of studies in which test bias was found, Schmidt (1988) and Hunter & Schmidt (1991) argued that many findings of bias were in fact attributable to low statistical power, restriction of range, and unreliability of measures.

Tests may be designed for reasons other than predicting an external criterion. For example, achievement tests instead focus on assessing current level of academic attainment. For this reason and because of the problems associated with the external methods approach to assessing test bias, internal methods have gained considerable popularity in recent years, and are more commonly associated with the term differential item functioning described above. In this method, the total test score or other items in the test are used as the criterion for examining group differences. Internal test bias occurs when the items in relationship to the other items on the test exhibit different measurement properties between two groups. Thus, the intent is to differentiate between true group differences and measurement bias (Camilli & Shepherd, 1994). Internal methods can be based on classical test theory (CTT), such as differential factor structure, item difficulties, or item discrimination, or based on item response theory (IRT). CTT or IRT may be used as the method for estimating either the level of ability used as the matching variable, or the item performance at each level of ability, some combination or both (Potenza & Dorans, 1995; Penfield & Lam, 2000)

<u>Observed Score Methods for Assessing Differential Item Functioning</u>

Many common internal methods for assessing DIF are based on classical test theory (CTT) using observed scores as the matching variable (Crocker & Algina, 1986). The CTT-based Mantel-Haenszel chi-square statistic, for example, is one of the most widely used bias detection methods (Milsap & Everson, 1993). Additional CTT-based methods for assessing DIF include the Delta-Plot method (Angoff, 1972; Angoff & Ford, 1973), chi-square methods (Scheuneman, 1979), the Standardization procedure (Dorans, 1989), and the logistic regression technique (Swamination & Rogers, 1990).

Many methods for detecting DIF were proposed primarily for dichotomous (right/wrong) items. As the use of polytomously scored items, those with more than two score categories such as those with multiple point rating scales, became more popular, many methods of DIF were extended to these items. Whereas, DIF in dichotomously scored items examines a differential probability of getting the item correct or incorrect, DIF in a polytomous item examines the differential probability of choosing each of a number of score categories. Recently, several methods based on observed scores have been extended to polytomous items. These include the Generalized Mantel-Haenszel (Zwick et al., 1993), Polytomous Standardization (Dorans & Schmitt, 1993), and the polytomous logistic regression technique (Rogers & Swamination (1993).

Mantel-Haenszel and Generalized Mantel-Haenszel Techniques

The Mantel-Haenszel x^2 procedure was introduced by Holland & Thayer (1988) based on a procedure used by Mantel and Haenszel in 1959. Similar to the x^2 method proposed by Scheuneman (1979), the Mantel-Haenszel x^2 procedure matches examinees on total test score. However, this procedure provides both a significance test as well as a measure of the effect size.

Mantel-Haenszel x^2 procedure utilizes a 2x2 contingency table that is created that shows the pass-fail status on one axis and the two groups of interest on the other matched at multiple intervals across the range of scores. The 2x2 contingency table is rewritten using probabilities, resulting in the following table:

		Score on item i		Total
		1	0	
Group	R	pRi	qRi	1
	F	pFi	qFi	1

This table is the replicated for *M* levels on the matching variable, such as total test score. The null hypothesis that no DIF exists, for the Mantel-Haenszel method is expressed as:

H_0: pRi = pFi

That is, the odds at getting any item right are equal for both the group R and group F.

Mantel-Haenszel (1959) developed a chi-square test of this null hypothesis (that no DIF exists) against an alternative hypothesis, known as the common odds ratio:

$$H_1 : \frac{\text{pRi}}{\text{pRi}} = \alpha \frac{\text{pFi}}{\text{pFi}} \qquad \text{i} = 1, \ldots. M$$

The parameter α is referred to as the common odds ratio in the *M* 2x2 tables. Under H_1, the value of α is the odds ration for all *M*:

$$\alpha = \frac{p_{Ri} q_{Fi}}{p_{Fi} q_{Ri}}$$

The Mantel-Haenszel (MH) chi-square test of the null hypothesis takes the following form:

$$MH - \chi^2 = \left[\left| \sum_m R_{rm} - \sum_m E(R_{rm}) \right| - .5 \right]^2 / \sum_m Var(R_{rm}),$$

where $E(R_{rm}) = E(R_{rm} | \alpha = 1) = N_{rm} R_{tm} / N_{tm}$,

$$Var(R_{rm}) = Var(R_{rm} | \alpha = 1$$

$$= \left[N_{rm} R_{tm} N_{fm} W_{tm} \right] / \left[N_{tm}^2 (N_{tm} - 1) \right]$$

Like the delta-plot method and full chi-square method, the Mantel-Haenszel procedure provides indexes of differential item difficulty but not of differential item discrimination. Despite this limitation, the Mantel-Haenszel procedure is extremely popular in testing due to its ease of calculation, use of the familiar chi-square statistic, and low cost (Millsap and Everson, 1993).

The Mantel-Haenszel procedure has also been extended for use in assessing DIF in polytomous items (Somes, 1986). Rather than comparing item scores for matched individuals, the generalized Mantel-Haenszel compares the vectors for item responses for matched individuals. Separate group-by-score contingency tables are created for each level of the matching variable and a chi-square test is used to asses the extent to which the distribution across the score categories differs for the groups. A significant difference in the distribution of the score categories indicates DIF. An advantage of this model is that it may retain the nominal nature of the rating categories while several other models assume an interval level measurement (Penfield & Lam, 2000).

Standardization and Polytomous Standardization Methods

The Standardization procedure matches examinees by total test score using all available appropriate data. Nonparametric item test regressions are estimated for both the reference, *r*, and focal, *f*, groups such that $E_f(I|M)$ defines the item test regression for the reference group and $E_r(I|M)$ defines the item test regression for the focal group where *I* is the item score variable and *M* is the matching variable. The null hypothesis is such that the item test regression for the reference group is equal to the item test regression for the focal group:

$$H_0: E_f(I|M) = E_r(I|M)$$

Any differences in item performance between the reference and focal groups matched on the measured attribute cannot be explained by the attribute. The item test regressions and differences in item performance across the groups matched on attribute are plotted and examined visually.

Although the item test regressions and difference plots provide visual representation of DIF, a simpler numerical index is required quicker identification of problematic items when large numbers of items are under examination. The Standardization method provides this numerical index as the Standardized p-difference. The Standardized p-difference (STD P-DIF) is defined as:

$$\text{STD P-DIF} = \sum_{m} N_{fm}(P_{fm} - P_{rm}) / \sum_{m} N_{fm}$$

Where:

N_{fm} = the number of examinees at m in the focal group

P_{fm} = the proportion correct, number of examinees who answer correctly over total number of examinees in the focal group

P_{rm} = the proportion correct, number of examinees who answer correctly over total number of examinees in the reference group

STD P-DIF ranges from –1 to +1 where positive values indicate the item favors the focal group and values of -.05 to .05 are considered negligible. Although both the Mantel-Haenszel procedure and the Standardization procedure utilize the 2(groups)-by-2(itemscore)-by-*M*(score level) contingency tables, Standardization focuses on differences in conditional proportions correct, rather than conditional odds ratios.

Dorans and Schmitt (1991) extended the standardization method to polytomous items. In Polytomous Standardization, the emphasis is on expected item score at each level rather than proportion correct. The comparisons of the item-test regression are amongst each score level weighted by the relative frequency of focal group members at that score level. Thus, the distinction becomes one of number of levels (proportion correct when right/wrong versus proportion at each score level).

Logistic Regression

Swamination & Rogers (1990) presented the logistic regression technique as a log-linear procedure that examines the interaction between group membership and ability level as an indicator of DIF. An extension of the Mantel-Haenszel procedure, this method is able to detect uniform as well as non-uniform DIF. The logistic regression model for DIF is presented below:

$$P(u=1) = \frac{e^z}{1+e^z}$$

where

$$z = \tau_0 + \tau_1\theta + \tau_2 g + \tau_3(\theta g).$$

In this equation, θ represents the observed trait, such as the total test score and *g* represents group membership. Hence, θg represents the interaction of group membership and the observed trait level. The coefficients, represented by τ, are estimated and tested to determine which are significantly different from 0. The coefficients τ_2 and τ_3 have significance in terms of the type of DIF displayed. An item is said to show uniform DIF if $\tau_2 \neq 0$ and $\tau_3 = 0$. An item shows nonuniform DIF if $\tau_3 \neq 0$ regardless of if $\tau_2 = 0$.

Rogers & Swamination (1993) found this procedure to be as powerful as the Mantel-Haenszel procedure at detecting uniform DIF and superior to the Mantel-Haenszel procedure at detecting non-uniform DIF. This procedure can be extended to polytomous data through the use of pairwise comparisons between score categories or combinations of score categories.

One issue associated with the use of procedures such as Mantel-Haenszel, Standardization, and Logistic Regression is the circular use of the overall test score as the matching variable. Dorans & Holland (1993) argue that this is acceptable if the test has demonstrated equal validity for both the reference and focal groups. There has also been disagreement over whether items that are found to function differently for two groups (i.e. show DIF) should be included as a part of the overall criterion or test score when the DIF analysis occurs. Holland & Thayer (1988) argued for the inclusion of the item.

SIBTEST

The basis of the simultaneous item bias test (SIBTEST) proposed by Shealy and Stout (1993) is that each item measures nuisance traits in addition to the main construct of interest. The focus is on how the performance of the referent and focal groups is affected by these nuisance traits. SIBTEST examines DIF at both the item and testlet level. This nonparametric procedure compares the original item responses of the reference and focal group using their scores on a matching subtest, consisting of items known to be unbiased as determined by a visual inspection or statistics. When one cannot identify a matching subtest of known unbiased items, then the procedure must be performed successively for $i = 1$ to n items, where n is the number of items on the total test, i is the item under investigation, and the matching subtest contains the remaining $n - 1$ items.

The SIBTEST has been found to be as powerful as the Mantel-Haenszel procedure in detecting uniform DIF when the ability distributions of the reference and focal group are the same and more powerful when the distributions are unequal (Narayanan & Swaminathan, 1994). A modified SIBTEST has been developed which is sensitive to nonuniform DIF and has also been extended to polytomous and multidimensional data (Chang, Mazzeo, & Roussos, 1996).

Item Response Theory

Although the use of classical test theory has been predominant in the evaluation of the psychometric properties of personality measures, the use of modern or item response theory, IRT, has increased (Reise & Hansen, 2003). Prior to discussing IRT-based methods for assessing differential item functioning, a brief overview of IRT is warranted.

IRT links the item response to the underlying trait, which is presumably measured by the scale of interest (Drasgow & Hulin, 1990). The plot of this relationship is provided by the Item Characteristic Curve (ICC), with the latent trait, or theta (θ), provided on the x-axis, and the y-axis representing the probability of a correct response. In the case of personality measures, the y-axis represents the probability of providing a response in the scaled direction as there is no "correct" response. As the amount of the trait increases so does the chance of obtaining a correct response. The ICCs can be summed to form the Test Characteristic Curve (TCC), so that an evaluation can be made at the test level.

The ICC is described by multiple IRT models. Both unidimensional and multidimensional models are available, as well as models used for both dichotomous and polytomous scored data. Some of the most commonly used models include the one-, two-, and three-parameters models for dichotomous data, and the Samejima's (1969) graded response model for polytomous data. These models vary in their assumptions and number of parameters used to describe the ICC (Ellis & Mead, 2002).

The three-parameter model uses the *a*-, *b*-, and *c*- parameter to describe the ICC. The *b* parameter is the difficulty parameter, the location on the θ scale where exactly half of the individuals taking the test will get the item correct or endorse the item in the keyed direction For items that are scored correct or incorrect as the value of *b* increases, the higher the level of ability required for the individual to have a 50% chance of getting the item correct. For personality measures, this means that as the value of the *b* parameter increases the greater the amount of the trait the individual must have before having a 50% chance of answering the item in the scaled direction. The second component in the three-parameter models is the *a* parameter, the item discrimination parameter. In the 3-parameter model, *a* is equal to the slope of the ICC at the point of maximum discrimination The final parameter in the three-parameter IRT is *c,* "pseudo-chance" or "pseudo-guessing" parameter. In personality measures, this parameter is interpreted as the likelihood that the respondent will endorse an item in the opposite direction of their true score (Zumbo, Pope, Watson, & Hubley, 1997). The equation for the ICC under the 3-parameter model is:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

where P_i (θ) = the probability that a randomly chosen examinee with ability θ answers item *i* correctly,

a_i = the item discrimination parameter, or the slope of the ICC at the point b_i on the ability scale,

b_i = the item difficulty parameter, or the point on the ability scale where the probability of a correct response is .5,

D = scaling factor equal to 1.7 to equate the logistic normal ogive functions.

Samejima's graded response model (1969) extends IRT to polytomous scales with ordinal response options, for example, as in a Likert-type scale or multiple point grading scale. The graded response model is a generalized form of the two-parameter model for dichotomous data and is appropriate for items where a respondent chose one of a number of ordered categories. For an item with m+1 categories, m difficulty parameter values are estimated for each item as well as one common discrimination parameter. The following equation is used to calculate the probability of an examinee replying to an item in a particular category or above, or the boundary response function (BRF):

$$P(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

The probability of responding to a single category requires subtracting the adjacent boundary from the cumulative probability, giving the item category response function (CRF). This is calculated as:

$$P_{ik}(\theta) = P_{i(k-1)}(\theta) - P_{ik}(\theta)$$

Using the item category response function, the expected item score for item i for examinee s can be calculated as using the following equation:

$$ES_{si} = \sum_{k=1}^{m} P_{ik}(\theta) X_{ik},$$

where

X_{ik} = the score or weight for category k,

m = the number of categories,

P_{ik} = the probability of responding to category k, defined above.

The first and last categories, which lack an adjacent boundary, are defined as $P_{i0}(\theta) = 1$ and $P_{m}(\theta) = 0$.

This model provides a particular advantage in the examination of personality measures as most major personality assessments feature polytomous response options, such as Likert-type scales (Zickar, 2002).

Prior to using IRT, the two basic assumptions of unidimensionality and local independence must be met. Although similar concepts, unidimensionality refers to the concept that only one trait or ability is measured by the items and local independence means that responses to test items are statistically independent, or uncorrelated. The assumption of unidimensionality is considered met when there is a “dominate” factor in the measure (Hambleton, Swaminathan, & Rogers, 1991). As they are related concepts, when the assumption of unidimensionality is met, so is the assumption of local independence. Provided these assumptions are met, IRT provides item characteristics, which are not sample dependent and estimates of ability which are not test dependent, unlike CTT.

IRT-Based Methods for Assessing Differential Item Functioning

Using IRT, DIF indicates that different groups of individuals have different item response functions and hence different item parameters. That is, on a valid test individuals from different subgroups with identical thetas should have equal probabilities of choosing a particular response. If they do not, then the item is said to be differentially functioning. In 1980, Lord provided the IRT definition of DIF as:

“If each test item in a test had exactly the same item response function in every group, then people of the same ability or skill would have exactly the same chance of getting the item right, regardless of their group membership. Such a test would be completely unbiased. If on the other hand, an item has a different item response function for one group than for another, it is clear that the item is biased.” (p. 212).

In general, DIF can be assessed using IRT through two general approaches: comparing the item parameters, which characterize the response functions, of two groups (Lord, 1980; Thissen, Steinberg, & Wainer, 1993) or examining the area between the ICCs (Hastins & Wardrop, 1981; Kim & Cohen, 1991; Linn, Levine, & Raju, 1988, 1990).

Several studies that have compared different procedures have found these two approaches provide similar results (Cohen & Kim, 1993, 1995; Raju, Drasgow, & Slinde, 1993; Shepherd, Camilli, & Averill, 1981).

DIF can also be classified as uniform (or consistent) DIF or nonuniform (or inconsistent) DIF. Uniform DIF occurs when the difficulty parameters are not equivalent across groups but the discrimination parameters are. When graphed, the ICCs for both groups do not cross, thus one group has an advantage over the other group at all levels of theta. In nonuniform DIF neither the discrimination nor difficult parameters are equivalent across groups. Here, the ICCs for the two groups cross at a point of theta. This indicates that the item is not related to the latent trait in the same way at all levels of theta across groups. When this occurs, this type of positive and negative DIF may cancel each other out. For this reason, area measures have been developed separating out both the signed (SA) and unsigned (UA) areas.

As with DIF procedures based on classical test statistics, several IRT-based methods have been extended to polytomous items. For example, Cohen, Kim and Baker (1993) extended Lord's chi-square and signed area indexes to Samejima's graded response model. Raju et al.'s DFIT framework extends the assessment of DIF not just to polytomous items but to the test level as well.

<u>Area Measures</u>

Methods relying on the area between the ICCs are descriptive in the effects of DIF. Examining the area between two ICCs presents a very visual demonstration of how different two groups are in their responses. Items with larger differences between the two ICCs function much differently for the two groups than items with smaller areas between the ICCs.

The area between the curves can be calculated over either the open (i.e. exact) interval or closed interval where only the area within limits that are typically those at the extremes of the scale, such as –4 and +4. Kim & Cohen (1991) found the exact and closed interval methods to provide similar levels of detection of DIF.

The equations for both the signed (SA) and unsigned (UA) area measures are provided by Raju (1988) as follows:

$$\text{SA} = (1-c)(b_F - b_R)$$

$$\text{UA} = (1-c)\left|\frac{2(a_F - a_R)}{Da_R a_F}\ln\left\{1+\exp\left[\frac{Da_R a_F(b_F - b_R)}{a_F - a_R}\right]\right\} - (b_F - b_R)\right|$$

In both equations, *a*, *b*, and *c* are the item parameters and *D* is the scaling constant to equate the logistic ogive to the normal ogive as described previously. Difficulties in estimating the *c* parameter can be overcome by estimating a common, or fixed, *c*. When c is not fixed and $c_F \neq c_R$, the exact area is infinite (Raju, 1998).

Significance tests are recommended instead of arbitrary cut-offs for area measures (Raju, 1990). When significance tests are applied, the signed and unsigned area measures provide results similar to other DIF detection methods (Raju, Drasgow, & Slinde, 1993). Cohen et al. (1993) extended Raju's area measures to the graded response model. For these items, DIF is examined as the area between the characteristic curves for each category for the reference and focal groups.

<u>IRT Models Using Parameter Differences</u>

When DIF is assessed through the item parameters the item response model is fitted to the data for both the focal and reference groups. As the item characteristic curve is based on the item parameters, the difference between the item parameters for the two groups is then examined. If this difference is found to be significant then DIF is said to exist (Theissen, Steinberg, & Wainer, 1993). These methods of assessing DIF may take into account differences on the *a*, *b*, or *c* parameter, or some combination. For example, for a given item the *a* parameters could be equal for the referent and focal group while the *b* parameters are different, making the item easier for one group than the other but equally discriminating.

Lord's Model

Lord (1977, 1980) introduced the assessment of DIF through the examination of differences in estimated item parameters for the reference and focal groups using a statistical test based on chi-square. The Lord statistic is expressed as:

$$\chi^2 - LORD = V'S^{-1}V.$$

In the above equation, V is the vector of the differences between the parameters for the reference group and the parameters for the focal group for the same item. S^{-1} is the inverse of the matrix of asymptotic variance-covariance of the vectors of differences between the parameters.

Results using this method has been found to correlate with the CTT-based chi-square and Delta Plot method (Linn & Werts, 1997). As previously mentioned, this method has been extended to Samejima's graded response model for polytomous items (Cohen et al., 1993). Under the graded response model, the vector V of differences between the item parameter estimates is given by:

$$V = \left(a_R - a_F, b_{1R} - b_{1F}, b_{2R} - b_{2F,}, ..., b_{(m,1)R} - b_{(m-1)F}\right)$$

and S^{-1} is the associated variance-covariance matrix. In this case, the null hypothesis follows a χ^2 distribution with *m* degrees of freedom.

Likelihood Ratio Test

The likelihood ratio (LR) test investigates the same null hypothesis defined by Lord that the item parameters do not differ between groups. In this method, the model is first simultaneously fitted to the data for both the reference and focal groups with "anchor" items constrained to have the same parameters for both groups. There are no such constraints placed on the items under investigation. Then, $G_1^2 = -2(\text{loglikelihood})$ is computed for the maximum likelihood parameter estimates. The model is then refitted under the constraint that the parameters are equal for the two groups for the items under investigation and $G_2^2 = -2(\text{loglikelihood})$ is computed. Then the likelihood ratio test $G^2(1) = -2(\text{loglikelihood})$ is used to examine the significance of the difference between the parameter of the reference and focal group.

Although the LR test requires two separate calibrations for each comparison for the parameters, it does not require equating of the reference and focal group parameters prior to the DIF analysis as the anchor items define a common metric. Both Lord's method and the likelihood ratio method test the null hypothesis that the item parameters for the two groups do not differ and both rely on the normality of the likelihood when p-values are obtained from the chi-square distribution. However, Lord's method further relies on the accurate estimate of the corresponding covariance matrix, (Σ), a difficult task. When the likelihood is normal and the covariance matrix is correctly estimated, Lord's method and the likelihood ratio method should perform equally well (Thissen, Steinberg, & Wainer, 1988).

<u>The DFIT Framework</u>

In 1995, Raju, van der Linden, & Fleer proposed a new IRT-based parametric procedure to examine differential functioning in dichotomous or polytomous scored data called differential function of items and tests (DFIT). Additionally, this procedure can also be used with multidimensional as well as unidimensional data and examines DIF at both the test and item-level. An additional benefit to the DFIT framework over other IRT based measures of DIF is its inclusion of multiple measures of DIF- compensatory (CDIF) and noncompensatory DIF (NCDIF).

One clear distinction between the DFIT framework and many other methods of detecting differential functioning is its ability to detect differential functioning at the test level as well as the item level as differential functioning can occur at the item level or test level, or both. This is an important distinction as decisions are made at the test level rather than at the item level. A combination of multiple items containing item bias can create bias at the test level (Shealy & Stout, 1993a). Small amounts of test bias present in multiple items can amplify to create test bias at the overall test level. However, previous research by Roznowski & Reith (1999) has found that item level bias does not necessarily lead to poor measurement quality at the scale level. For example, nontrait-related variance may cause an item to display DIF but no single source of nontrait-related variance is present in all the items. Additionally, item bias can cancel each other out when the errors are in opposite direction (Shealy & Stout, 1993). For example, the effect of an item that contains bias against males can be counteracted by the inclusion of an item that contains bias against females.

Prior to the 1990's (Shealy & Stout, 1993b) there were no measures for differential functioning at the test level. The DFIT framework provides an alternative IRT based parametric approach to assessing differential test functioning (DTF) with the additional benefits of allowing the summation of individual compensatory DIF (CDIF) values in the definition of DTF. That is, item level differential functioning is considered as additive, combining to test level differential functioning. This item level functioning is seen as adding to or subtracting from the test level differentially functioning in a compensatory manner. Two items with each absolute values of DIF but opposing signs will result in a zero net effect on the magnitude of DTF as the DIF of the two items cancel each other out at the test level. Because of the CDIF index in the DTF framework, it is possible to observe substantial item level differential functioning without observing significant test level differential functioning.

Unlike the CDIF index in DFIT, the non-compensatory DIF (NCDIF) index does not allow for summation of values. Similar to Lord's (1990) chi-square and Raju's area measures, the NCDIF index does not consider that items can compensate for bias in other items as it is assumed that other items are free of DIF except for the one under examination.

In IRT, $P_i(\theta_s)$ represents the probability that an examinee chosen at random with a given theta (θ) score answers item *i* in the scaled direction. For the DFIT framework, the one-, two-, or three- parameter IRT model can be used resulting in item parameters *a*, *b*, and *c* for the two groups- the reference (R) and focal (F) groups with item parameters estimated separately and assumed to be on the same scale for the two groups. If the item functions differently for the two groups, $P_i(\theta_s)$ will be different. For a polytomous item, this concept is the *ES*, or expected item score, previously described.

This concept is extended to the scale level using the expected proportion correct (EPC), also called the "true" score, and expressed as:

$$T_s = \sum_{i=1}^{n} P_i(\theta_s)$$

For each individual, two true scores are calculated- one as a member of the focal group (T_{sF}), the other as a member of the reference group (T_{sR}). A test is said to function differently if T_{sF} does not equal T_{sR}. The larger the difference between the two true scores the greater the extent of differential functioning of the test. DF for an examinee at the test level is then defined by:

$$D_s^2 = (T_{sF} - T_{sR})^2$$

Taken across examinees for the focal group (E), the average is written as:

$$DTF = E_F D_s^2 = E_F (T_{sF} - T_{sR})^2$$

Which is equivalent to:

$$DTF = \sigma_D^2 + (\mu_{TF} \mu_{TR})^2 = \sigma_D^2 \mu_D^2$$

where σ_D^2 is the variance of *D*, μ_{TF} is the mean true score for the focal group, and μ_{TR} is the mean true score for the same focal group members as if they were members of the reference group.

According to Raju et al. (1995) the compensatory DIF index is defined by the following equation:

$$CDIF_i = \in (d_i, D) = Cov(d_i, D) + \mu_{d_i} \mu_D$$

Cov(d,*D*) in the above equation is the covariance between the difference in item probabilities for the item *i* (d_i) and the difference between the two true scores. As differential functioning at the test level is the sum of compensatory differential functioning at item level, DTF is also defined as:

$$DTF = \sum_{i=1}^{n} CDIF_i$$

Flowers, Oshima, & Raju (1999) demonstrated the application of the DFIT framework to polytomous data using Samejima's (1969) graded response model. The DFIT framework for polytomous items only differs from dichotomous in calculation of the item true score, using the *ES* defined previously for polytomous items.

DFIT Significance Tests

Chi-square (χ^2) significance tests for the DFIT framework were also described by Raju et al. (1995). Assuming the difference between EPCs (*D*) is normally distributed with a mean μ_D and a standard deviation of σ_D each examinee's score be transformed into a Z score:

$$Z_s = \frac{D_s - \mu_D}{\sigma_D}$$

The chi-square test for DTF is:

$$\chi^2_{N_F} = \frac{N_F(DTF)}{\hat{\sigma}_D}$$

A significant χ^2 value indicates that CDIF is present in one or more items. Raju et al. (1995) recommend that any items containing CDIF be identified and removed with χ^2 recalculated with the remaining items. This process should continue until χ^2 is no longer significant. Because of the manner in which CDIF works in the DTF framework, there is no associated significance test for CDIF. The identification of DTF indicates that CDIF is present.

Significant tests have, however, been developed for the NCDIF index using a similar chi-square test as for DTF. The chi-square test for NCDIF is as follows:

$$\chi^2_{N_F} = \frac{N_F(NCDIF)}{\hat{\sigma}_D}$$

Because previous research (Fleer, 1993) has found that χ^2 tests of significance for the DFIT framework is overly sensitive to large sample sizes, Raju et al. (1995) has recommended setting a cutoff value of NCDIF in addition to the significant χ^2 test. Although a cutoff value of 0.016 has been used in research using the DFIT framework in both dichotomous (Fleer, 1993) and polytomous items (Flowers, 1995; Searcy, 1998), a cutoff value of 0.096 is now recommended for use with polytomous data with five response options (Swander, 1999; Raju, 2000) although it has been argued that this cutoff value is too set high if the measure under investigation is used for selection purposes (Meade & Lautenschlager, 2005; Meade, Lautenschlager, & Johnson, 2005).

Gender Differences in Personality

Gender differences have frequently been found on numerous personality measures that examine neuroticism and similar traits such as negative affectivity, depression, and anxiety (Costa, Terracciano, & McCrae, 2001; Gomà-i-Freixanet, Valero, Puntí Zuckerman, 2004; Goodwin & Gotlibb, 2004; Jorm, 1987; Lee & Ashton, 2004; Lynn & Martin, 1997; Reynolds, 1998; Santor, Ramsey, & Zirroff, 1994). Such gender differences have been found across a wide variety of ages (Jorm, 1997; Weiss, Costa, Karuza, Duberstein, Friedman, & McCrae, 2005) and in multiple cultures (Abdel-Khalek, & Alansari, 2004; Brachman & Costello, 1963; Costa, Terracciano, & McCrae, 2001; Lynn & Martin, 1997). In addition, gender differences have been noted in anxiety-related disorders and clinical depression, with which there is a noted link to neuroticism and its associated facets (Aneshensel, Rutter, & Lachenbruch, 1991; Jorm, 1987).

In addition to individual studies noting a gender difference, several meta-analyses have also been conducted. Jorm (1987) conducted a meta-analysis focusing on gender differences in neuroticism. The primary interest of this study was in neuroticism as it relates as a risk factor for depression, which tends to show gender differences in rates of occurrence during the middle period of life. Jorm found a mean effect size of 0.32 based on 26 relevant studies. Gender differences in neuroticism were found at all ages, although an inverted -U trend was found such that the differences decreased in childhood and later in life.

In 1994, Feingold conducted two meta-analyses on gender differences in personality and found only minor differences on the trait of anxiety, a facet of neuroticism, for men versus women. The first of the two meta-analyses, using studies previously reviewed by Maccoby and Jacklin in 1974, found in one that females scored higher than males on measures of anxiety (weighted mean d = -.29). In the other meta-analysis, a replication of an earlier meta-analysis by Hall (1984), the unweighted mean d was found to be -.15, again with females scoring higher.

Gender differences have been noted on specific personality measures, as well. Using data from the NEO-RI-R collected between 1992 and 2000, Costa et al. (2001) revisited the results of Feingold (1994) and looked at gender differences on the Big Five scale level. Costa et al. found an average effect size for adults in the United States of .51 for neuroticism with women scoring higher than men. Additionally, at the facet level, mean z-score differences (d) between men and women were highest for the facets of Vulnerability (d=.44) and Anxiety (d=.40).

Gender differences have been perhaps most intensely examined in regards to the Eysenck neuroticism scales such as the Eysenck Personality Questionnaire (Lynn and Martin, 1997; Francis, 1993a; Zuckerman, Joireman, Kraft, & Kuhlman, 1999) and the Eysenck Personality Questionnaire- Revised Abbreviated (Forrest, Lewis, Shevlin, 2000; Shevlin, Bailey, and Adamson, 2002). Lynn and Martin (1997) for example, found that women scored higher than men in neuroticism in samples drawn from 37 different countries. Zuckerman et al. (1999) found significant gender differences in the Neuroticism scale of the Eysenck Personality Questionnaire in a sample of college students. Forrest et al. (2000) also found a statistically significant association between scores on the Neuroticism scale of the Eysenck Personality Questionnaire- Revised Abbreviated and gender in terms of regression effects, with females scoring higher on the measure, although the effect sizes found were small. Additionally, Francis (1993b) has argued that there are two components within Eysenck's Neuroticism scales, one that is sex-related, or highly correlated with gender, and another component which is sex-free, or uncorrelated with gender. She found women scored higher than men on the sex-related component but not the sex-free component.

Age Differences in Personality

In addition to gender differences in personality, age differences in personality have also been examined. Across personality measures, research has found that neuroticism scores generally decrease with age (Loehlin & Martin, 2001; Terracciano, McCrae, Brandt & Costa, 2005; Costa, McCrae, Zonderamn, Barbano, Lebowitz, & Larson, 1986; Roberts, Walton, & Viechtbauer, 2006; Costa, Herbst, McCrae, & Seigler, 2000). Such an effect has been found on both the Eysenck personality scales (Loehlin & Martin, 2001) and the NEO-PI-R (Terracciano, McCrae, Brandt & Costa, 2005).

In fact, the research on the NEO-PI-R has proven to be so compelling that the inventory authors have revised their previous statements that "personality is stable after age 30" (Costa, & McCrae, 1988) and "individual differences in personality traits, which show at least some continuity from early childhood on, are also essentially fixed by age 30" (McCrae & Costa, 1994, pg 173). The inventory authors, after reviewing their own research as well as that of others, have recently agreed with the conclusion that scores on the Neuroticism scale decline with age (Costa & McCrae, 2006). Current research tends to support the idea that neuroticism decreases until approximately age 80, when it become stable or increases slightly (Terracciano et al., 2005). Other research has noted only slight decreases between ages 55 to 80 (Steunenberg, Twisk, Beekman, Deeg, & Kerkhof, 2005; Weiss, Costa, Karuza, Duberstein, Friedman, & McCrae, 2005). This decrease in scores on the Neuroticism scale has also been found in other countries or cultures (McCrae, Costa, Lima, Simões, Ostendorf, Angleitner, Marusic, Bratko, Caprara, Barbaranelli, & Chae, 1999; Steunenberg, 2005; McCrae, Costa, Hrebickova, Urabanek, Martin, Oryol, Rukavisknikov, & Senin, 2004).

Research has also examined the possibility that changes in personality as one ages may be moderated by gender. However, the research on this interaction has been mixed. Some studies have found an interaction between gender and age. For example, Srivastava, John, Gosling, and Potter (2003) found that Neuroticism decreased in a sample of ages 21-60 for women but not men. However, other research examining age differences moderated by gender found no significant affect of age by gender interactions (Terracciano, McCrae, Brandt & Costa, 2005). And one meta-analysis (Roberts, Walton, Viechtbauer, 2006) concluded differences in Emotional Stability, the inverse of Neuroticism, between the two genders were not significant as one aged.

<u>True Differences or Measurement?</u>

However, what these group differences mean and why they exist is still a puzzle. Do group differences on personality measures indicated true group differences, bias in measurement, or a combination of both? Biological theories have been proposed to explain gender differences in personality. These theories point to innate differences between the sexes that have developed through evolution as each sex has developed characteristics that have helped them adapt and ensure their survival. Additional biological theories have pointed out hormonal differences between the genders and their effects on personality and mood states. For example, the greater tendency of women toward behaviors such as crying is well documented (e.g. Lombardo, Crester, Lombardo, & Mathis, 1983). Although research support for gender differences in neuroticism from this perspective is unavailable, one can look at other related traits and disorders, such as depression. Specific research has addressed gender differences found in incidences of depression from a biological theory perspective. One in-depth review found that support for these biological theories of differences in the incidences of depression was inconclusive (Nolen-Hoeksema, 1987).

Gender differences in personality may also be examined through social role theory (Eagly, 1987; Eagly & Wood, 1999), which suggests that differences in personality are developed as both genders adopt a shared expectation of appropriate conduct for men and women. Shevlin et al. (2002) has argued that the basis for gender differences found in neuroticism scales may be based not in biological sex differences, but in these learned gender roles. Although research support for this theoretical basis for the differences found in neuroticism is limited, some support was found by Krampen, Effertz, Jostock, and Muller (1990). Krampen et al. (1990) found that sex-role orientation produced larger effect sizes and explained more of the variance in a measure of neuroticism (German Eysenck Personality Inventory) than did physical sex (male/female). However, this research is largely limited by small sample size.

That gender and age differences found on scales of Neuroticism are found may also be an interaction between true differences and bias in measurement. As described by Jorm (1987), one explanation is that neuroticism presents itself differently in males versus females. For example, females high in neuroticism may be more prone to cry while males loose their temper. A scale that contains more items that are oriented toward females, such as one that includes more items about the intensity or frequency of crying behavior, would show a bias due to the selection of the items.

Recent Research Examining DIF in Personality Measures

To explore if the differences found are true differences or a result of the measurement properties at the item level, research examining differential item functioning in popular personality measures has appeared in recent literature, particularly examining differential item functioning by gender. Analyses of differential item functioning by gender have been conducted at least one subscale from the Multidimensional Personality Questionnaire (Smith and Reise, 1999), the NEO-PI-R (e.g. Reise, Smith & Furr, 2001), the Eysenck Personality Questionnaire Revised — Abbreviated (Shevlin, Bailey, and Adamson, 2002), the Spanish version of the Eysenck Personality Questionnaire (Escorial, & Navas, 2007), the Hogan Personality Inventory (Sheppard, Han, Colarelli, Dai & King, 2006) and the Beck Depression Inventory (Santor et al.., 1994). Because of our interest in potential gender bias in Neuroticism scales, only research examining this trait and related scales is visited here.

In 1999, Smith and Reise explored DIF on the 23-item Stress Reaction scale of the Multidimensional Personality Questionnaire (MPQ). Mean differences were found between men and women on the scale and several items were found that demonstrated significant levels of DIF that served to raise women's scores. However, several items were also found that contained DIF that served to raise men's scores. The authors note a pattern in the items displaying DIF in that items that were easier for women to endorse tended to be grouped into a content factor centered around emotional vulnerability and help seeking responses to stress. Items that were found to be easier for men to endorse dealt with the experience of tension, moodiness, and irritation. This suggests that the different genders may experience different manifestations of stress and that the balance of items tapping each construct was necessary to avoid test level differential functioning. As suggested by Smith (2002), such research provides insight into constructs that may have otherwise been unnoticed or ignored.

In 2001, Reise, Smith, and Furr followed-up on the research presented by Smith and Reise (1999) and examined DIF on the facet scales of the NEO-PI-R. Descriptive statistics found women scoring significantly higher than men on four of the six facet scales of the NEO-PI-R Neuroticism scale, with the Anxiety facet demonstrating the largest mean difference. Using the generalized partial credit model, Reise et al. examined the facet scales of the NEO-PI-R Neuroticism scale for potential DIF. Of the six facet scales, the largest amount of items found displaying DIF, and the largest amount of DIF, belonged to the Anxiety facet scale.

In this facet, both items that displayed significant DIF serving to raise men's scores and items displaying significant DIF serving to raise women's scores were identified. The effects of the DIF were found to cancel each other out when considered at the scale level as test characteristic curves calculated separately for both gender groups were found to be nearly overlapping. However, the authors also note the finding that contained within the Anxiety facet are two identifiable content factors- one tapping fearfulness and the other tapping tension and worry with mean differences found primarily in the fearfulness factor. This finding of multidimensionality provides insight into the source of DIF in the Anxiety facet.

Research on the Beck Depression Inventory (Santor et al., 1994) did not find gender bias in the majority of items on the scale, despite gender differences in mean scores. Of the three items that display moderate amounts of DIF the greatest amount of item bias was found on an item relating to “distortion of body image” with women responding more strongly in both clinically depress and college samples. Bias was not addressed at the test level.

In his unpublished dissertation, Shepperd (1997) examined gender DIF in the Hogan Personality Inventory. In his analysis of the Adjustment scale, the scale of the HPI found to correspond with Emotional Stability/Neuroticism in the Big Five, the author found that most, 18 out of 25 analyzed items, contained no DIF in either direction. Four items were found to contain DIF in favor of males, three in favor of females. Of the four items found to contain DIF in favor of males, three dealt specifically with the individual’s relationship with their parents. One item measuring anxiety was found to contain DIF. The item “I often feel anxious” found DIF favoring females. No mean level score information was reported.

Missing from the current wave of DIF/DTF research is research examining differential functioning by age in personality measures. Also missing is research on the International Personality Item Pool (http://ipip.ori.org/), a new public domain personality item pool generating substantial interest and research. The International Personality Item Pool (IPIP) consists of a large pool of items, which have been combined to form public-domain measures corresponding to current popular personality measure. Of particular interest for the current research is the 60-item scale developed to correspond to the NEO-PI-R Neuroticism scale. Mean gender differences have been reported on IPIP scales measuring the NEO-PI-R Neuroticism (Buchanan, Johnson, & Goldberg, 2005; Gow, Whiteman, Pattie, & Deary, 2005). Additionally, differences in mean scores have also been found across age ranges with older populations of both genders scoring higher in Emotional Stability, hence, lower in Neuroticism (Gow, Whiteman, Patty, & Dreary, 2005). Given the gender and age differences have been found on the IPIP, as well as other similar measures, and that differential functioning of items in other similar measures has been found, it is important to examine differential functioning in the IPIP. This study proposes to add to the current literature by extending the current DIF research to this new measure.

Proposed Study

This study is proposed to examine potential item- and scale-level differential functioning on the Neuroticism scale of a newer personality measure, the International Personality Item Pool scale designed to measure the NEO-PI-R constructs, using the item response theory based DFIT framework. Based on substantial research indicating gender differences in scores on various scales measuring the trait of neuroticism, including the IPIP, it is expected that this sample will also display gender differences. Thus, the following is hypothesized:

Hypothesis 1: Mean gender differences will be found on the Neuroticism scale scores with women scoring higher than men.

As the IPIP-NEO measure is designed to measure the same constructs as the NEO-PI-R, it is expected that similar results to that of Reise, Smith, and Furr (2001) will be found here. Given that Reise et al., and other authors examining differential item functioning in scales measuring personality constructs related to neuroticism, found differential item functioning in items, and that these items are similar in nature to those found on the IPIP-NEO, it is expected that at least some of the items on the Neuroticism scale of the IPIP-NEO will display DIF. However, it can also be hypothesized that, as with the NEO-PI-R, the IPIP-NEO will consist of both items that are easier for women to endorse and of items that are easier for men to endorse. These items will essentially serve to cancel each other out, leading to no finding of differential functioning at the test level. Thus, it is hypothesized that:

Hypothesis 2a: The Neuroticism scale of the IPIP will demonstrate significant differential functioning at the item level.

Hypothesis 2b: The Neuroticism scale of the IPIP will not demonstrate significant differential functioning at the scale level.

Based on the previous research described above, it is also expected that the greatest number of items displaying DIF will occur in the Anxiety facet. Not only is this consistent with Reise, Smith, and Furr (2001), but also with Shepperd (1997) who found the specific item “I often feel anxious” displayed DIF. Two hypotheses are presented regarding the Anxiety facet of the IPIP-NEO:

Hypothesis 3a: The Anxiety facet scale of the Neuroticism scale of the IPIP will contain the greatest number of items demonstrating differential item functioning when compared to the other facet scales.

Hypothesis 3b: The Anxiety facet scale of the Neuroticism scale of the IPIP will contain the greatest amount of differential item functioning when compared to the other facet scales.

Reise, Smith, and Furr (2001) found that of the items displaying DIF serving to raise women’s scores on the Anxiety facet, two specifically addressed fear issues. It can be argued from a gender role theoretical standpoint, that these items are easier for women to endorse, as it is more socially acceptable, and thus a part of their socially constructed gender role, for women to acknowledge and demonstrate fears. Research has found gender differences in self-reported level of fear (e.g. Dillon, Wolf, & Katz, 1985). Additional research specifically investigating confidence in expressing fear found women were significantly more confident in expressing feelings of fear (Blier and Blier-Wilson, 1989), Therefore, the following hypothesis is presented:

Hypothesis 4: Fear items on the Anxiety facet scale will be easier for women to endorse and will display DIF. Specifically, these items are "Fear for the worst" and "Am afraid of many things."

No research has yet examined personality measures for DIF by age. However, given that age differences have been found on scales measuring Neuroticism, it is suspected that at least some items will functioning differently for different age groups but that at the scale level, these differences will cancel out so that no difference will be found at the scale level.

Hypothesis 5a: The Neuroticism scale of the IPIP would demonstrate significant differential functioning by age at the item level in at least a subset of the items.

Hypothesis 5b: that the Neuroticism scale of the IPIP will not demonstrate significant differential functioning by age at the scale level.

Method

Participants

Substantial numbers of respondents are required for DIF analysis. For example, sample sizes less than 1000 for both the referent and focal group have been found to impair the performance of NCDIF and CDIF (Searcy, 1998). For this reason, the largest available sample of respondents was used for this study. Currently, this sample is a dataset collected by John A. Johnson and consists of over 20,000 respondents.

23,994 respondents anonymously completed the 300-item IPIP measures via the internet between August 6, 1999 and March 19, 2000. This dataset was previously examined for invalid response protocols. Protocols were discarded if they showed evidence of duplicate responses, inattentive responding, greater than average missing responses, and unacceptably low levels of consistency. This resulted in a final sample of 20,933 respondents. Complete information on the method used for eliminating protocols is available from Johnson (2005).

The resulting sample consisted of 7743 (36.9%) males and 13249 (63.1%) females. All individuals reported ages between the ages of 10 and 99. However, less than 1% of the respondents reported ages over 58 years of age. This figure is consistent with what has been found in other studies utilizing Internet samples (Gosling, Vazire, Srivastava, & John, 2004). Because of the limited number of individuals providing data over age 58 and the research providing evidence that personality stabilizes in later years, individuals over age 55 were not used in the analysis examining differential functioning by age group. Instead the sample was divided into groups as follows: under age 20 (n=7333), ages 20 to 30 (n=8080), over 30 to 55 (n=5226).

Research examining web based assessment using the IPIP-NEO measure has found a similar factor structure as well as virtually identical scale intercorrelations and reliabilities, compared to the paper and pencil version (Buchanan, Johnson, & Goldberg, 2005). Additional previous research (Trippe & Harvey, 2002) reported similar item response functions between an IPIP-NEO dataset collected via the Internet and a IPIP-NEO data set collected via paper and pencil.

<u>Measure</u>

International Personality Item Pool (Goldberg, 1999). The International Personality Item Pool consists of 1412 items developed by Lewis R. Goldberg, in conjunction with researchers in the Netherlands and Germany. From these items scales have been developed to measure the Big Five domains and the Five Factor Model (FFM) domains (as measured by the NEO-PI-R), as well as a variety of other scales. These scales have been administered to samples along with the commercially available scales (including the NEO-PI-R) in order to ascertain their similarity in measurement constructs. Research investigating the psychometric properties of scales developed from the IPIP has been supportive and encouraging of their use (Lim & Ployhart, 2006; Guenole & Chernyshekno, 2005). The scales consisting of IPIP items are public domain and are available for download at www.ipip.ori.org.

For this study, the scale paralleling the FFM construct of Neuroticism as measured by the subscale of the NEO-PI-R (Costa & McCrae, 1992) was used. Each item in the scale consists of a personality descriptive term with a 5-point likert-type scale with responses ranging from very inaccurate to very accurate. Approximately half of these items were inversely coded so that agreeing with the item indicated lower level of Neuroticism and were reverse coded for data analysis.

There are 60 items on this scale with 10 items corresponding to each of the subscales of the NEO-PI-R. On IPIP scale, these subscales are labeled Anxiety, Anger (labeled Anger Hostility on the NEO-PI-R), Depression, Self-Consciousness, Immoderation (labeled Impulsiveness on the NEO-PI-R) and Vulnerability. The coefficient alphas for the six Neuroticism subscales measured by the IPIP scale measuring the NEO-PI-R Neuroticism subscales range from .77 to .88. These scales have been found to correlate .72 to .80 with the NEO-PR-R Neuroticism subscales (Goldberg, 2000). The items of the Neuroticism scale of the International Personality Item Pool are provided in Table 1 in Appendix A.

Unidimensionality of Scale

To test for unidimensionality of the data, exploratory factor analytic (EFA) procedures were employed, specifically principal components analysis (PCA). Both the scree plot and the percent of variance accounted for by the first factor were examined to test if the data met unidimensionality assumptions for application of an item response theory model.

The first factor accounted for 27.93% of the variance. In order to ensure acceptable item calibration, the percent of variance accounted for by the dominant factor must be at least 20 percent (Reckase, 1979). Thus, the results of the PCA demonstrated that the criterion of greater than 20% of the variance accounted for was met. The results of the PCA are reported in Table 2 in Appendix B.

In determining unidimensionality of the subscale, the first factor must be clearly separated from the other factors on the scree plot in order to show the presence of one dominant first factor. The scree plot, provided in Figure 1 in Appendix B, was also examined for a clear separation of the first factor from the other factors as illustration of one dominant first factor. As shown, there is one factor that is more dominant than the others. The ratio of the first factor to the second factor is approximately 4:1.

One could argue the dimensionality of the data based on the fact that more than one factor would have been retained if the eigenvalue greater than one criteria had been used. In fact, if this criterion were to be used, nine factors would be retained. However, based on the results of Reckase (1979) and Drasgow & Parsons (1983, 1982) the amount of variance accounted for by the first factor found here is sufficient to apply an IRT model to examine the properties of the underlying general factor.

Additionally, the correlations between the scores of the subscales is provided in Table 3 and the item correlation matrix is provided in Table 4.

IRT Parameter Estimation

Samejima's (1969) graded response model was used for item parameter estimation, resulting in one discrimination parameter and four difficulty parameters for each item. The computer program MULTILOG (Thissen, 1995) was used for parameter estimation. Parameters were calibrated separately for the reference and focal groups for each comparison. The parameters, one *a* (discrimination) and four *b* (difficulty) parameters, for the gender groups are provided in Tables 5 and 6 in Appendix C and for the age groups in Tables 7, 8, and 9 in Appendix C.

Equating

As item parameters calibrated in separate MULTILOG runs are on different metrics, a transformation was required to put the groups on the same metric in order for the groups to be compared. This was accomplished using the Fortran-based computer program EQUATE (Baker, 1993), which employs Stocking & Lord's (1993) iterative test characteristic curve procedure. This characteristic curve method minimizes differences across groups in the score intervals in the transformation of the scale and has been shown to provide improved bias detection over non-iterative linking methods in both dichotomous models (Candell & Drasgow, 1998) and polytomous models (Hildago-Montesinos & Lopez-Pina, 2002). EQUATE provided the linear transformation coefficients used in the DFIT program to transform the reference group parameters onto the metric group, allowing group comparisons.

As the equating process is not considered accurate if DIF items are used in the equating process, the DIF analysis was performed after items were identified as having

DIF, and the scales re-equated with the DIF items taken out. For the gender analysis, this required three iterations as item 6 was initially flagged for DIF following the first analysis but not flagged for DIF after removed from the equating. As an additional item 15, was flagged during the second iteration, a third iteration was conducted with item 15 removed from the equating and item 6 replaced into the equating procedure. This third iteration confirmed item 15 was correctly identified as DIF and item 6 was not. The final equating constants are provided in Appendix C following the item parameters.

DIF Analysis

Since differential functioning at both the item and overall test level of the polytomous measure was of interest, the DFIT framework (Raju, van der Linden, & Fleer, 1993) was used to assess differential item functioning. DFIT provides additional advantages over other models in that it can detect uniform and nonuniform DIF and has been found to show decreased false positive rates over other DIF detection methods (Bolt, 2002).

As the larger group in sample size, the females were designated as the reference group and the males were designated as the focal group for comparison purposes. For the age comparison, the middle age group, consisting of respondents between the ages of 20 and 30, was the designated as the reference group for comparison against both the under 20 group and the over the age of 30 to age 55 group. Hence, the middle age group was compared against the older group as the focal group and then again against the younger group as the focal group.

The DFIT5 program (Raju, 2003) was used to estimate the NCDIF, CDIF and DFIT indexes. The number of theta estimates was reduced to 3000 as the DFIT program will only allow a maximum of 3000 to be used (Flowers, personal communication). However, 3000 has been found to be sufficient for stable estimates.

NCDIF is considered statistically significant with the NCDIF index less than or equal to 0.096 and chi-square is statistically significant at .01. The DTF cutoff value is this same level of significance multiplied by the number of items in the scale and is outputted by the computer program DFIT (Raju, 2003). If the results indicate that DTF is significant an iterative process will occur of removing the item with the largest DIF and re-examining DTF until DTF is no longer significant

Results

Mean Differences

Mean scores for the Neuroticism scale and the six subscales are provided in Appendix B for the two genders, along with other descriptive statistics. Statistically significant mean differences between genders were found in the Neuroticism scale ($t(20990)= -28.127$, ($p<.05$)) with women scoring higher (M= 180.956, SD=36.503) than men (M=166.222, SD= 36.820). As Levene's statistic for equality of variances was not significant ($p>.05$), equal variances were assumed.

Statistically significant mean differences between genders were also found in the subscales of the Neuroticism scale. For the Anxiety subscale, statistical differences were found ($t(20990)= -35.343$, ($p<.05$)) with women scoring higher (M=31.979, SD=7.606) than men (M=27.9241, SD= 7.754). As Levene's statistic for equality of variances was not significant ($p>.05$), equal variances were assumed.

Statistical differences were found for the Anger subscale as well ($t(15734.337)= -19.621$, ($p<.05$)). Women scored higher (M=30.358, SD=8.914) than men (M=27.979, SD= 9.239). As Levene's statistic for equality of variances was significant (*p*>.05), equal variances were not assumed.

For the Depression subscale, statistical differences were found (*t*(15910.956)=-10.345, (*p*<.05)) with women scoring higher (M=28.074, SD=9.333) than men (M=26.674, SD= 9.543). As Levene's statistic for equality of variances was significant (*p*>.05), equal variances were not assumed.

The comparison of the Self-consciousness scale found significant differences (*t*(20990)=-14.526, (*p*<.05)) with women again scoring higher (M=29.976, SD=7.607) than men (M=28.399, SD=7.572). As Levene's statistic for equality of variances was not significant (*p*>.05), equal variances were assumed.

Scores on the Immoderation subscale was also higher for women (M=33.392, SD=6.942) then men (M=31.736, SD=6.791) with statistical differences (*t*(16491.141)= -16.907, (*p*<.05)). As Levene's statistic for equality of variances was significant (*p*>.05), equal variances were not assumed.

Lastly, for the Vulnerability subscale, statistical differences were found (*t*(16463.306)=-35.221, (*p*<.05)) with women scoring higher (M=27.361, SD=7.368) than men (M=23.694, SD= 7.223). As Levene's statistic for equality of variances was significant (*p*>.05), equal variances were not assumed.

However, as these statistical differences may be due largely to sample size and may not be practically significant, a measure of effect size, Cohen's *d* was also computed and is also reported in Appendix D. Using the cut-offs described by Cohen (1992), medium effects were found for the overall Neuroticism scales, the Anxiety subscale, and the Vulnerability subscale. These effect sizes were consistent with those reported by Cohen et al. (2001). Thus, Hypothesis 1 received support.

Gender Comparisons

It was hypothesized that the Neuroticism scale of the IPIP would demonstrate significant differential functioning by gender at the item level, and this hypothesis was supported. In fact, six items, 10% of the items on the 60 item measure, did display NCDIF. These items are: Item 15 (Am often down in the dumps): NCDIF= .105 χ= 82429 ($p<.01$), Item 16 (Find it difficult to approach others): NCDIF=.124 χ= 301559 ($p<.01$), Item 20 (Am often in a bad mood): NCDIF=.148 χ= 149318 ($p<.01$), Item 28 (Only feel comfortable with friends): NCDIF= .119 χ>999999($p<.01$), Item 33 (Feel desperate): NCDIF=.201 χ= 64966 ($p<.01$), and Item 39 (Feel that my life lacks direction): NCDIF=.119 χ= 239165 ($p<.01$). NCDIF values for all items are located in Appendix E.

Hypothesis 2b, that the Neuroticism scale of the IPIP will not demonstrate significant differential functioning by gender at the scale level, was also supported. The DTF index was below the cut-off value of 5.760 (DTF= 3.263, χ= 80909.71, $p<.01$). As DTF was not found, no items demonstrated significant CDIF. As the average CDIF value was .05439, and DTF is the sum of the CDIF index of the items, additional items demonstrating CDIF in the positive direction, or greater positive values of CDIF on the items already demonstrating CDIF, would have been needed to provide a finding of significant DTF.

Although the Anxiety subscale was hypothesized to have the greatest number and amount of differential item functioning, in fact, no items in that subscale demonstrated significant NCDIF. Instead, three items, half of those identified with NCDIF, were in the Depression subscale. These items are Item 15: “Am often down in the dumps,” Item 33: “Feel desperate,” and Item 34: “Feel that my life lacks direction.” Thus, Hypotheses 3a and 3b did not receive support.

Hypothesis 4, fear items on the Anxiety facet scale will be easier for women to endorse and will display DIF, was not supported. The items "Fear for the worst" and "Am afraid of many things" did not display significant NCDIF.

Age Comparisons

Hypothesis 5a, that the Neuroticism scale of the IPIP would demonstrate significant differential functioning by age at the item level, was supported. In the mid/young group comparison, NCDIF was displayed by two items, Item 11: NCDIF= .116 χ= 28515 ($p<.01$) and Item 53: NCDIF= .247 χ= ($p<.01$). These items are "Don't know why I do some of the things I do" and "Never spend more than I can afford." In the mid/older group comparison, only one item displayed NCDIF. This item was Item 24: NCDIF= .113 χ= 193635 ($p<.01$). This item is "Can't make up my mind."

Hypothesis 5b, that the Neuroticism scale of the IPIP will not demonstrate significant differential functioning at the scale level, was also supported. For the middle-age group and older age group comparison, the DTF index was well below the cut-off value of 5.760 (DTF= .19590, χ= 104081.90, $p<.01$). For the middle- age group and younger age group comparison, the DTF index was also well below the cut-off value of 5.760 (DTF= .00910, χ= 3621.62, $p<.01$).

Discussion

The results of the comparisons by gender are first discussed, followed by the comparisons by age. An overall discussion of the results follows the item level comparison discussion. Limitations of the study and suggestions for future research are provided.

Gender Comparisons

Comparisons of the male and female gender groups found six items that displayed differential item functioning. These were items 15, 16, 20, 28, 33, and 39. Three of these items were from the six item Depression subscale. The category response curves and boundary response options for these items are graphically displayed in Appendices F and I. The response options for all items are 1) Very Inaccurate 2) Moderately Inaccurate 3) Neither Inaccurate nor Accurate 4) Moderately Accurate 5) Very Accurate. When administered some items were coded in opposite direction and reversed for data analysis.

Item 15 is "Am often down in the dumps." An examination of the difficulty parameters found that women (b_1= -1.100, b_2=-0.190, b_3=.436, b_4=1.42) were less likely to endorse the item than men (b_1=-1.376, b_2=-0.506, b_3=0.141, b_4=1.164) at the same level of theta. Respondents indicated endorsement of the item by selecting from response options ranging from "very inaccurate to "very accurate". The boundary response functions (BRFs) for this item are displayed graphically in Table 4 in Appendix F where each line shows the probability of the respondent selecting from the category or above. The category response functions (CRFs) for this item are displayed graphically in Table 13 in Appendix H. From the graphs and the item parameters, it can be seen how women would be -1.100 standard deviations below the mean theta to have a .5 probability of responding "very inaccurate" but that men would be -1.376 standard deviations below the mean of theta to have the same probability of responding in the same manner. A possible explanation for the finding of DIF in this item could be the interpretation of "down in the dumps." This idiomatic phrase could have been interpreted differently by women than men.

Item 16, "Find it difficult to approach others," was also endorsed less by women (b_1= -2.23, b_2= -0.423, b_3=.436, b_4=2.44) than men (b_1=-2.499, b_2=-0.976, b_3=-0.139,

b_4=1.664) at the same level of theta. As evident by the CRF in Figure 14 in Appendix H the item displays greater DIF at the "moderately accurate" and "very accurate" response categories. For example, females at a theta value of 3 have a .70 probability while males have a .91 probability of selecting the response option "Very Accurate". In comparison, at a theta value of -3 females have a .76 probability while males have a .70 probability of selecting the response option "Very Inaccurate". Thus, it is evident that at the upper level of theta men are more willing to endorse this item as descriptive of their personality than women are despite less difference in willingness to endorse opposing response options at the emotional stability side of the scale.

Figure 15 in Appendix H demonstrates the CRFs for Item 20, "Am often in a bad mood" where DIF is evident across the range of θ where males were more likely to endorse response options in the direction of indicating higher levels of Neuroticism. This difference is also evident from examining the expected true scores where at a .5 theta value level of the trait, men are expected to have an item score of 3.212 and women an item score of 2.724. This finding is consistent with the finding of Smith & Reise (1991) that men found items dealing with moodiness on the Multidimensional Personality Questionnaire easier to endorse.

Item 28 is "Only feel comfortable with friends" where men were again more likely (b_1=-3.54, b_2=-1.397, b_3=-.376, b_4=1.593) to select response options agreeing with the statement than women (b_1=-3.1, b_2=-.736, b_3=.261, b_4=2.34). As with item 16, this item addresses social interaction and displays a pattern of greater differences at the "Very Accurate" than "Very Inaccurate" response options.

Men were also more likely to endorse Item 33 which states "Feel desperate." The pattern of DIF is shown in Figure 8 in Appendix F and Figure 17 in Appendix H.

Differences in the two groups are very evident at higher levels of theta (indicating higher levels of the trait of neuroticism). For example, at a theta value of 2, females have a .37 probability of indicating that this statement very accurately depicts them while men have a .75 probability of choosing that same level of agreement. The largest NCDIF index from the gender comparisons resulted from this particular item. This indicates that this item had the largest difference between the subgroup item parameters and largest expected true score difference between males and females. An analysis of the IRT-based expected item true scores shows a .48 difference in the expected true scores at a theta value of 1 with males having an expected true score of 3.907 and females an expected true score of 3.426 at this theta value. This is displayed graphically in Figure 22 in Appendix J. The average absolute difference in the expected item raw scores of the two groups is .448, which would be the square root the NCDIF index.

Men were also more likely to agree with the statement "Feel that my life lacks direction", which is Item 39. As evident by examination of the graphs and item parameters, smaller differences were noting in respondents that selected the "Very Inaccurate" response option than in respondents selecting the other options, suggesting that while both genders at the Emotional Stability end of the scale may have similar tendency to disagree with the statement that their life lacks direction, at higher levels of neuroticism, men are more likely to agree with the statement.

Despite all items identified with NCDIF favoring men, at the overall test level the Neuroticism scale did not display DTF by gender. This finding is possible due to the nature of the summation of CDIF, a signed compensatory measure, rather than NCDIF, an unsigned noncompensatory measure, to examine differential functioning at the test level. Unlike NCDIF, CDFIT takes into account the interaction of all items on the test and is used to identify items that should be removed had significant DTF been found (Raju et al., 1995). While the measure contains six items that were found to function differently at a level that reached significant levels, at the overall scale level low level differences of item functioning in the other direction in other items serve to "cancel out" the items that had DIF reaching levels of significance, providing comparable theta estimates based on the entire scale. As there was no finding of DTF, no items were identified for removal from the scale with CDIF.

Age Comparisons

Only three items were found to function differently between different age groups. Two of the items displayed DIF between the under 20 and the 20-to-30 age group. Only one item displayed DIF between the 20-to-30 age and over 30-to-55 age group.

Item 11 is "Don't know why I do some of the things I do." A demonstration of the category response curves and item parameters show substantial differences in the *b* parameters between the under the age of 20 group (b_1=-4.176, b_2=-2.369, b_3=-1.151, b_4=1.075) and the age 20 to 30 group (b_1=-2.77, b_2=-1.35, b_3=-.454, b_4=1.41). These differences are more pronounced particularly amongst those with a lower level of the trait of Neuroticism. Only at the uppermost levels of the Neuroticism scale are the two groups of equal or similar probability of endorsing the response options. At the lower levels of Neuroticism, the Emotional Stability end of the scale, the under the age of 20 group has a lower probability of endorsing the option that this item does not accurately describe them. This item also shows substantial differences in the *a* parameter, demonstrating different levels of discrimination for the different subgroups. This is demonstrated in the boundary response functions, shown in Table 10 in Appendix G, where one can see that one of the lines indicating a response cross.

Item 53, "Never spend more than I can afford," also displayed DIF between the aged 20-to-30 and under-20 group. While the item discriminated similarly between the two groups (age 20-to-30 a=.335; under-20 a=.328), there were differences in the b-parameters where the item was more likely to be endorsed by the 20-to-30 age group (b_1=-6.08, b_2=-2.28, b_3=-0.974, b_4=3.08) than the under-20 group (b_1=-4.195, b_2=-0.492, b_3=1.217, b_4=6.002). This item demonstrated large true score differences between the two groups. This is graphically represented in Table 23 in Appendix J. As individuals tend to have different financial responsibility as one grows older, it is not unsurprising that this item would demonstrate differential functioning by age. This item may have interpreted differently by the two groups in the context of different financial responsibilities occurring as responsibility of financial burden tends to shift from a parental responsibility to that of the individual during this age period.

The only item to display DIF between the 20-to-30 and the over 30-to-55 group was Item 24, which is "Can't make up my mind." As demonstrated by the b parameters, the middle age group (b_1=-2.34, b_2=-0.639, b_3=.157, b_4=1.97) was more likely to endorse the item in the keyed direction at the same trait level as the older group (b_1=-1.706, b_2=-.156, b_3=0.636, b_4=2.365). One potential explanation could be the interpretation of this item as admitting indecision and the differential levels of social desirability of admitting such as one grows older.

At the test level, as with the gender analysis, the Neuroticism scale did not display DTF for either age group comparison. Thus, no items were identified for removal for the scale with CDIF. Had a finding of DTF been found, the measure should not have been used until items with CDIF had been eliminated.

Overall Discussion

The finding of DIF in items itself, while an interesting finding, should be considered only an initial step in any deliberation on potential scale revisions. DIF is a statistical property of the item and should be only one step contributing to the judgmental decision of an item being biased for or against a subgroup. The decision to replace items demonstrating DIF must be carefully considered. A careful evaluation of the construct should be made as any replacements of items could, in essence, serve to change the construct being measured.

DIF is one measure of the psychometric quality. When scales are undergoing revisions, if items are already performing poorly for other reasons the scale developer may want to consider replacing the item with a similar item of higher quality that does not function differently between subgroups. One such example would be item 53 which poorly discriminated between those high and low on the trait, contributing little to the evaluation at the scale level. A potential replacement item could be one which shows greater discrimination along the range of the scale as well as functions similarly for the subgroups of interest while still measuring the construct of interest.

However, a dilemma may then occur if a potential replacement item is determined to show substantial discrimination and a good measure of the construct but demonstrates DIF by subgroup. The scale developer must take into account the subgroups that function differently and weigh it against the impact on the measurement of the construct. Some subgroup comparisons, such as gender or ethnic DIF, may be a more relevant determination than other comparisons, such as DIF between chocolate-lovers versus chocolate-haters. Similarly, the developer must also consider that removing items demonstrating DIF may ultimately narrow the construct being measured and removing items with DIF may remove measurement of sub areas of the construct of interest. This is particularly relevant to the Depression subscale where half of the items demonstrated DIF by gender group. Replacing all of these items could significantly change the construct being measured.

Another consideration the scale developer should take into account is the impact of other items flagged for DIF. For example, while item 53 was flagged for DIF in the direction of making the item easier for the 20-to-30 group to endorse, item 11 is designed to measure the same subscale, was flagged for DIF in the opposite direction, and ultimately, when all items were considered, there was no differential functioning at the total scale level. Thus, removal of the items may not be necessary when evaluations are made at the scale, not item, level.

Although DIF was found in 10% of items, it was not found to provide a clear total test score advantage (or disadvantage as lower scores on Neuroticism scales tend to be more desirable). Although the DIF favored males on the gender analysis, men did not tend to score higher on the instrument and differential functioning at the test level was not found. However, it is still suggested that the measure be evaluated for why DIF is occurring and if replacing the items could eliminated DIF without impairing measurement of the construct.

Lastly, it should also be noted that while the measure examined in this research was the 60-item measure of the IPIP designed to assess the Neuroticism scale of the NEO-PI-R including the six subscale facets, there are also shorter measures created from a subset of the items on the longer 60-item scale. Two additional scales have been created- a ten and a twenty item scale. These scales are publicly available on the IPIP website at: http://ipip.ori.org/newNEOKey.htm#Neuroticism. These scales do not purport to include any measurement of the subscales. Of the items included in the shorter scales, only one item was found to display DIF in this examination of the longer measure. This item, "Am often down in the dumps" displayed DIF by gender and is used in both the twenty and ten item scales. Therefore, particular attention should be paid to this item in any subsequential qualitative examination of the item.

Limitations

A limitation to the age group analysis is that there were not adequate numbers of respondents over the age of 55 to allow analyses using that age group. Although research suggests that personality stabilizes in later years, the limited sample size in this research did not allow empirical examinations of individual items that may function differently in older age ranges. Items that displayed differential functioning between two age groups may have been more pronounced when older age ranges were examined. Additionally, different selection of age range groups may have resulted in different parameter estimates leading to different DIF values. Age ranges for this study were selected based on sample size and early research suggesting that personality changes may not be as pronounced after the age of 30 (Costa, & McCrae, 1988). Without clear delineation of comparison groups (i.e. by gender or other dichotomous variables) the selection of comparison groups is somewhat arbitrary. Different delineation of age groups may have resulting in other or additional items identified as functioning differently.

The mode of original data collection may also be considered a limitation of the study. Although the anonymity of the internet may encourage more honest responses, the veracity of the responses to demographic data cannot be verified. Although the dataset was examined for invalid response protocols and responses were discarded if they showed evidence of duplicate responses, inattentive responding, greater than average missing responses, or unacceptably low levels of consistency, it is possible that some invalid or duplicated responses were included in the dataset as no restrictions on repeated responses from a single computer were in place during data collection.

Although internet data collection has the ability to increase some demographic variability (e.g. geographic location), it may restrict other demographic characteristics, such as socio-economic status (Gosling et al., 2004). Thus, while the data collection method facilitated the collection of a large number of responses, the respondents themselves may not be representative of the general population. In general, this is a limitation that many research studies struggle with. However, it should be noted that DIF studies on the IPIP collected via traditional paper-and-pencil methods may not identify the same items as differentially functioning as this study did. Previous studies have shown that men and women use the internet in different ways with women using the internet more for social purposes such as communicating through e-mail (e.g. Jackson, Ervin, Garder, & Schmitt, 2001) and such differences may have affected which items were identified with gender DIF. Potential differences between the groups of individuals who chose to access and complete the measure may have impacted the results, such as the finding of two items dealing with social interaction demonstrating DIF.

Future Research

Qualitative examination of the items by subject matter experts in the field of personality, gender, and aging to gain further insight into potential causes of the statistical identification of differential item functioning is suggested. Any explanation of potential differences made herein is post hoc explanation made by the researcher, without input of subject matter experts that may provide further insight. Such further exploration of the item content, and the construct the items deem to measure, would be a critical step in any potential item or scale revisions. Focus groups, cognitive interviews, and judgmental reviews by experts in personality assessment could help determine possible causes of the DIF, such as misinterpretation of items or the response scale, or differential cognitive processes used in responding to the item. A statistical finding of DIF in and of itself offers no explanations for why such differences are occurring.

Future research should expand analyses of differential item functioning to other IPIP scales designed to measures similar constructs. For example, IPIP scales have been developed to assess Anxiety, not just as defined by the NEO measure but also Apprehension as defined by the 16PF measures. As these measures are highly correlated, investigation into differential functioning between genders on this measure would be beneficial. Differential item functioning statistics should also be conducted to examine the measurement invariance of these measures on other demographic variables such as ethnicity and other age range groups.

In addition, it would be beneficial to the field of personality assessment and the understanding of sub-group differences in item interpretation to further examine items similar to those on the IPIP scales that measure similar constructs. For example, Item 15, "Am often down in the dumps" is similar to the items "Was down in the dumps" on the Center for Epidemiologic Studies Depression Scale (Radloff, 1977). However, one study found no DIF by gender for this item (Gelin & Zumbo, 2003). Further analyses and evaluation of this item to determine why men and women are interpreting "Was down in the dumps" similarly in the context of that Center for Epidemiologic Studies Depression Scale and differently to the item "Am often down in the dumps" in the context of the IPIP Neuroticism measure, could further help explore the construct and the interpretation of commonly used items to measure Depression and Neuroticism. Also as previously noted, this item is the only item identified with DIF in this study that is also included on the shorter 10- and 20- item scales of IPIP designed to measure the construct of Neuroticism.

It would also be beneficial to the investigation of the construct of neuroticism and gender differences in the interpretation of items to examine other measures. While there is no IPIP scale designed to measure the Eysenck Personality Questionnaire-Revised (EPQ-R), one recent study on the Spanish version found similar results to this study. When items measuring Neuroticism were investigated, they found mean differences with women scoring higher but all items identified as NCDIF using the DFIT methodology were in favor of men whereas no DTF at the scale level was found (Escorial & Navas, 2007). The authors do not include the items which had DIF but a comparison of the EPQ-R and the IPIP-NEO Neuroticism scale is warranted as a future avenue of research.

Additional research should be conducted to replicate the findings of this study on other samples, such as those collected via paper and pencil administrations. Replication of this study on a sample collected using traditional paper and pencil methods in particular could also strengthen the argument for the comparability of administration modes and would address one limitation of this study. Follow up studies examining DIF/DTF by mode of administration would also be possible on such a sample. While computerized administrations of personality and attitudinal measures have been found to be comparable to paper and pencil administrations (e.g. Pinsoneault, 1996; Whitaker, 2007) these studies did not offer the level of anonymity this administration did. It is possible that the anonymous mode of administration allowed individuals who are not normally prone to share feelings to researchers and assessors to be more expressive of their true self in responding to the questions on this scale.

Conclusion

As use of the public-domain International Personality Item Pool scales has increased over the past decade (Goldberg, et al., 2006) research on the psychometric properties of the items and scales have become more and more critical. This research adds to the current research literature examining the psychometric properties of the measure by examining one measure constructed from the IPIP for occurrences of differential item functioning by gender and three age ranges. While six items were found to display differential item functioning by gender and three items by age, no differential functioning at the test level was found in this study, suggesting that while measurement invariance occurs at the item level, at the total score level the instrument serves to measure the construct of Neuroticism similarly across gender and age groups. Items demonstrating DIF and implications for potential scale revision were discussed. Qualitative analysis of the items identified with a finding of statistical DIF for why the item is functioning differently between the subgroups is recommended. Future research should examine items from additional IPIP scales and similar items on other scales proposed to measure the construct of Neuroticism.

References

Abdel-Khalek, A. M. & Alansari, B. M. (2004). Gender differences in anxiety among undergraduates from ten Arab countries. Social Behavior & Personality, 32, 649-656.

Allen v. Alabama State Board of Education No. 81-697-N (consent decree filed with United States District Court for the Middle District of Alabama Northern Division, 1985).

Angoff, W. H. (1993). Perspective on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 3-23). Hillsdale, NJ: Lawrence Earlbaum.

Aneshensel, C.S., Rutter, C.M., & Lachenbruch, P.A. (1991). Social structure, stress, and mental health: Competing conceptual and analytic models. American Sociological Review, 56, 166-178.

Baker, F. B. (1995). EQUATE 2.1: Computer Program for Equating Two Metrics in Item Response Theory [Computer program]. Madison: University of Wisconsin, Laboratory of Experimental Design.

Bishop, Y.M., Fienberg, S.E., & Holland, P.W. (1975). Discrete multivariate analysis. Cambridge, MA: MIT Press.

Blier, M. J. & Blier-Wilson, L. A. (1989). Gender differences in self-rated emotional expressiveness. Sex Roles, 21, 287-295.

Bolt, D. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. Applied Measurement in Education, 15, 113-141.

Brachman, H.M., & Costello, C.G. (1963). Cultural and sex differences in extraversion and neuroticism reflected in responses to a children's personality inventory. British Journal of Educational Psychology, 32, 254-257.

Buchanan, T., Johnson, J. A., & Goldberg, L.R. (2005). Implementing a five-factor personality inventory for use on the internet. European Journal of Psychological Assessment, 2, 115–127.

Camilli, G., & Shephard, L. (1994). Methods for Identifying Biased Test Items. Thousand Oaks, CA: Sage Publications.

Chang, H., Mazzeo, J., & Roussos, L. (1994). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. Research Report.

Cleary, T.A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Cohen, A.S., Seock-ho, K., & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. Applied Psychological Measurement, 17, 335-350.

Costa, P.T. & Mc Crae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. Journal of Personality and Social Psychology, 54, 853-863.

Costa, P.T. & McCrae, R. R. (2006). Age changes in personality and their origins: Comment on Reobers, Walton, and Viechtbauer (2006). Psychological Bulletin, 132, 26-28.

Costa, P.T. & McCrae, R. R. (1992). Four ways the big five are basic. Personality and Individual Differences, 13, 653-665.

Costa, P.T., McCrae, R.R., Zonderman, A.B., Barbano, H.E., Lebowitz, B., & Larson, B.M. (1986). Cross-sectional studies of personality in a national sample: 2. stability in neuroticism, extraversion, and openness. Psychology and Aging, 1, 144-149.

Costa, P.T., Herbst, J.H., McCrae, R. R., & Seigler, I. C, (2000). Personality at midlife: Stability, intrinsic maturation and responses to life events. Assessment, 7, 365-378.

Costa, P.T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. Journal of Personality and Social Psychology, 81, 322-331.

Dillon, K. M., Wolf, E., & Katz, H. (1985). Sex roles, gender, and fear. The Journal of Psychology, 119, 355-359.

Dorans, N.J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. Applied Measurement in Education, 2, 217-233.

Dorans, N.J. & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In Holland, P. W. & Wainer, H. (Eds.) Differential Item Functioning. pp. 35-66. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Eagly, A. H. (1987). Sex Differences in Social Behavior: A Social-Role Interpretation. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Eagly, A. H. & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles, American Psychologist, 54, 408-423.

Ellis, B. B., & Mead, A. D. (2002). Item analysis: Theory and practice using classical and modern test theory. In Rogelberg, Steven G. (Ed). (2002). Handbook of research methods in industrial and organizational psychology. (pp. 324-343). Malden, MA, US: Blackwell Publishers.

Escorial, S. & Navas, M.J. (2007). Analysis of the gender variable in the Eysenck Personality Questionnaire- Revised Scales using differential item functioning techniques. Educational and Psychological Measurement, 67, 990-1001.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. Psychological Bulletin, 116, 429–456

Fleer, P. F. (1993). A Monte Carlo assessment of a new measure of item and test bias. Unpublished doctoral dissertation, Illinois Institute of Technology, IL.

Flowers, C.P., Oshima, T.C., & Raju, N.S. (1999). A description and demonstration of the polytomous-DFIT framework. Applied Psychological Measurement, 23, 309-326.

Forrest, S., Lewis, C.A., & Shevlin, M. (2000). Examining the factor structure and differential functioning of the Eysenck Personality Questionnaire revised-abbreviated. Personality and Individual Differences, 29, 579-588.

Francis, L. J. (1993a). Personality and religion among college students in the U.K. Personality & Individual Differences, 14, 619-622.

Francis, L. J. (1993b). The dual nature of the Eysenckian neuroticism scales: A question of sex differences? Personality and Individual Differences, 15, 43-59.

Furnham, A. & Mitchell, J. (1991). Personality, needs, social skills, and academic achievement: A longitudinal study. Personality and Individual Differences, 12, 1067-1073.

Furnham, A. & Zacherl, M. (1986). Personality and job satisfaction. Personality and Individual Differences, 7, 453-459.

Gelin & Zumbo, (2003) Differntial item functioning results may change depending on how an item is scored: An illustration with the center for epidemiological studies depression scale. Educational and Psychological Measurement, 63, 65-74

Godderd, R., Patton, W., Creed, P. (2004). The importance and place of neuroticism in predicting burnout in employment service case managers. Journal of Applied Social Psychology, 34, 282-296

Goldberg, L.R. (2000). International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality and Other Individual Differences [online]. Available: http://ipip.ori.org/ipip/.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C. R., Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. Journal of Research in Personality, 40, 84-96.

Gomà-i-Freixanet, M., Valero, S., Puntí, J., & Zuckerman, M. (2004). Psychometric Properties of the Zuckerman-Kuhlman Personality Questionnaire in a Spanish Sample. European Journal of Psychological Assessment, 20, 134-146.

Goodwin, R. D., & Gotlibb, I.H. (2004). Gender differences in depression: the role of personality factors. Psychiatry Research, 126, 135-142.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O.P. (2004). Should we trust web-based studies? American Psychologist, 59, 93-104.

Gow, A. J., Whiteman, M.C., Pattie, A. & Deary, I. J. (2005). Goldberg's 'IPIP' big-five factor markers: Internal consistency and concurrent validation in Scotland. Personality and Individual Differences, 39, 317-329.

Green, D.R. (1975). What does it mean to say a test is biased? Education and Urban Society, 8, 33-52.

Guenole, N. & Chernyshenko, O. S. (2005). The suitability of Goldberg's big five IPIP personality markers in New Zealand: A dimensionality, bias, and criterion validity evaluation. New Zealand Journal of Psychology, 34, 86-96.

Hambleton, R. K., Swaminathan, H., & Rogers, H.J. (1991). Fundamentals of Item Response Theory. Newbury Park: Sage

Hildago-Montesinos, M.D. & Lopez-Pina, J. A. (2002) Two-stage equating in differential item functioning detection under the graded response model with the Raju area measures and the Lord statistic. Educational and Psychological Measurement, 62, 32-44.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenzel procedure. In H. Wainer & H. Braun (Eds.) Test Validity (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Huang, C., Church, A.T., & Katigbak, M.S. (1997). Identifying cultural differences in items and traits: Differential function in the NEO Personality Inventory, Journal of Cross-Cultural Psychology, 28, 192-218.

Jackson, L.A., Ervin, K.S., Garder, P.D., & Schmitt, N. (2001). Gender and the internet: Women communicating and men searching. Sex Roles, 44, 363-376.

Jensen, A.R. (2000). Testing: The dilemma of group differences. Psychology, Public Policy, & Law, 6, 121-128.

Johnson, J. A. (2005). Ascertaining the validity of web-based personality inventories. Journal of Research in Personality, 39, 103-129.

Jorm, A.F. (1987). Sex differences in neuroticism: A quantitative synthesis of published research. Australian and New Zealand Journal of Psychiatry, 21, 501-506.

Kim. S. & Cohen, A. S. (1995). A comparison of Lord's chi-square, Raju's area measures, and the likelihood ratio test on detection of differential item functioning. Applied Measurement in Education, 8, 291-312.

Kirkaldy, B., Thome, E., & Thomas, W. (1989). Job satisfaction among psychosocial workers. Personality and Individual Differences, 10, 191-196.

Krampen, G. Effertz, B. Jostock, U., Muller B. (1990). Gender differences in personality: Biological and/or psychological? European Journal of Personality, 4, 303-317.

Lee, K. & Ashton, M.C (2004). Psychometric properties of the HEXACO personality inventory. Multivariate Behavioral Research, 39, 329-358.

Lim, B. & Ployhart, R. E. (2006). Assessing the convergent and discriminant validity of Goldbery's international personality item pool: a multitrait-multimethod examination. Organizational Research Methods, 9, 29-54.

Linn, R. L., Levine, M.V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. Applied Psychological Measurement, 5, 159-173.

Loehlin, J.C. & Martin, N.G. (2001), Age changes in personality traits and their heritabilities during the adult years: Evidence from Australian twin registry samples. Personality and Individual Differences, 30. 1147-1160.

Lord, F.M. (1977). A study of item bias using item characteristic theory. In Y.H. Pooringa (Ed.), Basic problems in cross-cultural research (pp. 19-29). Amsterdam: Swets & Zeilinger.

Lombardo, W. K., Crester, G. A., Lombardo, B., Mathis, S.C. (1983). Fer cryin' out loud- there is a sex difference. Sex Roles, 9, 987-995.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum.

Lynn, R. & Martin, T (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. Journal of Social Psychology, 13, 369-373.

Macmillan. (1975). Guidelines for creating positive sexual and racial images in educational materials. New York, NY: Macmillan.

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

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748.

McCrae, R. R., & Costa, P.T. (1994). The stability of personality: Observations and evaluations. Current Directions in Psychological Science, 3, 173-175.

McCrae, R. R., Costa, P. T., Lima, M. P., Simões, A., Ostendorf, F. Angleitner, A., Marusic, I., Bratko, D., Caprara, G. V., Barbaranelli, C., & Chae, J. (1999). Age differences in personality across the adult life span: Parallels in five cultures. Developmental Psychology, 35, 466-477.

McCrae, R.R., Costa, P.T., Hrebickova, M., Urabanek, T., Martin, T.A., Oryol, V.E., Rukavisknikov, A.A., & Senin, I.G. (2004). Age differences in personality traits across cultures: Self report and observer perspectives. European Journal of Personality, 18, 143-157.

McGraw-Hill (1968). Recommended guidelines for multiethnic publishing in McGraw-Hill Book Company publications. New York: McGraw Hill.

Meade, A. W. & Lautenschlager, G. J. (2005). Sensitivity of DFIT tests of measurement invariance for likert data. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Meade, A. W., Lautenschlager, G. J., & Johnson, E. C. (2007). A Monte-Carlo examination of the sensitivity of the DFIT framework for tests of measurement invariance with Likert data. Applied Psychological Measurement, 31, 430-255.

Milsap, R.E. & Everson, H.R. (1993). Methodology review: Statistical approaches for assessing measurement bias. Applied Psychological Measurement, 17, 297-334.

Narayanan P. & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. Applied Psychological Measurement, 18, 315-328.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. Psychological Bulletin, 101, 259-282.

Petersen, A.N.S. & Novick, M.R. (1976). An evaluation of some models for culture-fair selection. Journal of Educational Measurement, 13, 3-29.

Pinsoneault, T. B. (1996). Equivalency of computer-assisted and paper-and-pencil administered versions of the Minnesota Multiphasic Personality Invetory-2. Computers in Human Behavior, 12, 219-300.

Raju, N. S. (1988). The area between two item characteristic curves. Psychometrika, 53, 495-502.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. Applied Psychological Measurement. 14, 197-207.

Raju, N. S; Drasgow, F. & Slinde, J. A. (1993). An empirical comparison of the area methods, Lord's chi-square test, and the Mantel-Haenszel technique for assessing differential item functioning. Educational & Psychological Measurement, 53, 301-314.

Raju, N., van der Linden, W., & Fleer, P. (1995). IRT-based internal measures of differential functioning of items and tests. Applied Psychological Measurement, 19, 353-368.

Radloff, LS (1977). The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement 1, 385-401.

Ramsey, P. A. (1993). Sensitivity Review: The ETS Experience as a Case Study. In Holland, P. W. & Wainer, H. (Eds.) Differential Item Functioning. pp. 35-66. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rantanen, J. Pulkkinen, L. & Kinnunen, U. (2005). The big-five personality dimensions, work-family conflict, and psychological distress: A longitudinal view. Journal of Individual Differences, 26, 155-166.

Reckase, M. D. (1979). Unifactor Latent Trait Models Applied to Multi-Factor Tests: Results and Implications. Journal of Educational Statistics, 4, 207-230.

Reise, S.P. & Hansen, J. M. (2003). A discussion of modern versus traditional psychometrics, as applied to personality assessment scales. Journal of Personality Assessment, 81, 93-103.

Reise, S.P., Smith, L., & Furr, R.M. (2001). Invariance on the NEO-PI-R neuroticism scale. Multivariate Behavioral Research, 36, 83-110.

Reynolds, C.R. (1998). Need we measure anxiety differently for males and females? Journal of Personality Assessment, 70, 212-221.

Reynolds, C.R. (1982). The problem of bias in psychological assessment. In C.R. Reynolds, & T.B. Gutkin (Eds.), The handbook of school psychology (pp. 178-208), New York: John Wiley.

Roberts, B. W., Walton, K. E., &Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. Psychological Bulletin, 132, 1-25.

Robie, C., Zickar, M., & Schmit, M.K.J., (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. Human Performance, 14, 187-207.

Rogers, H. J. & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedure for detecting differential item functioning. Applied Psychological Measurement, 17, 105-116.

Roznowski, M. & Reith, J. (1999). Examining the measurement quality of tests containing differentially functioning items: Do biased items result in poor measurement? Educational & Psychological Measurement, 59, 248-269.

Santor, D. A., Ramsay, J. O., & Zuroff, D. C. (1994). Nonparametric item analyses of the Beck Depression Inventory: Evaluating gender item bias and response option weights, Psychological Assessment, 6, 255-270.

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. Psychometrika Monograph Supplement, 34 (4, Pt. 2).

Scheuneman, J.D. (1975). A method of assessing bias in test items. Journal of Educational Measurement, 16, 143-152.

Searcy, C. A. (1998). A Monte-Carlo investigation of the DFIT framework applied to polytomous data under the graded response model. Unpublished doctoral dissertation, University of Georgia, Athens, GA.

Shealy, R, T. & Stout, W.F. (1993a) An item response theory model for test bias. In Holland, Paul W. (Ed); Wainer, Howard (Ed). (1993). Differential item functioning. (pp. 337-347). Hillsdale, NJ, US: Lawrence Erlbaum Associates.

Shealy, RJ. & Stout, W.F. (1993b). A model-based standardization approach that separates true bias/DIF from ability group differences and detects test bias/DTF as well as item bias/DIF. Psychometrika, 58, 159-194.

Shepperd Jr, R.L. (1997). Differential Item Functioning in the Hogan Personality Inventory. Unpublished dissertation, Central Michigan University.

Sheppard, R., Han, K., Colarelli, S.M., Dai , G. & King, D.W. (2006). Differential Item Functioning by Sex and Race in the Hogan Personality Inventory. Assessment, 13, 442-453.

Shevlin, M., Bailey, F., & Adamson, G. (2002). Examining the factor structure and sources of differential functioning of the Eysenck Personality Questionnaire Revised – Abbreviated. Personality and Individual Differences, 32, 479-487.

Sireci, S.G. & Geisinger, K.F. (1998). Equity issues in employment testing. In Sandoval, J. H. & Frisby, Craig L. (Eds), Test interpretation and diversity: Achieving equity in assessment. (pp. 105-140). Washington, DC, US: American Psychological Association.

Smith, L.L. (2002). On the usefulness of item bias analysis to personality psychology. Personality and Social Psychology Bulletin, 28, 754-763.

Smith, L. L. & Reise, S.P. (1998). Gender differences on negative affectivity; An IRT study of differential item functioning on the multidimensional personality questionnaire stress reaction scale. Journal of Personality and Social Psychology, 75, 1350-1362.

Somes, G. W. (1986). The generalized Mantel-Haenszel statistic. The American Statistician, 40, 106-108.

Srivastava, John, O.P. Gosling, and Potter (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? Journal of Personality and Social Psychology, 84, 1041-1053.

Stark, S., Chernyshenko, O.S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important? Journal of Applied Psychology, 89, 497–508.

Steunenberg, B., Twisk, J. W. R., Beekman, A. T . F. (2005). Stability and change of neuroticism in aging. Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 60, 27-33.

Swamination, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27, 361-370.

Swander, C. (1999). Assessing the differential item functioning of items and tests of a polytomous employee attitude survey. Unpublished masters thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Terracciano, A. McCrae, R.R., Brandt, L. J., & Costa, P. T. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore longitudinal study of aging. Psychology and Aging, 20, 493-506.

Terris, W. (1997). The traditional regression model for measuring test bias is incorrect and biased against minorities. Journal of Business & Psychology, 12, 25-37.

Theissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In Wainer, H & Braun, H.I. (Eds). (1988). Test Validity. (pp. 147-169). Hillsdale, NJ, US: Lawrence Erlbaum Associates.

Theissen, D., Steinberg, L., Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In Holland, Paul W. & Wainer, H. (Eds). Differential item functioning. (pp. 337-347). Hillsdale, NJ, US: Lawrence Erlbaum Associates.

Trippe, M. & Harvey, R.J. (2002). Item response theory analysis of the IPIP big-five scales. Paper presented at the 18th annual conference for the Society of Industrial Organizational Psychology, Orlando, FL.

Weiss, A., Costa, P. T., Karuza, J., Duberstein, P. R., Friedman, B., & McCrae, R.R. (2005). Cross-sectional age differences in personality among medicare patients aged 65 to 100. Psychology and Aging, 20, 182–185.

Whitaker, B. G. (2007). Internet-based attitude assessment: does gender affect measurement equivalence? Computers in Human Behavior, 23, 1183-1194.

Zickar, M. J. (2002). Modeling data with polytomous item response theory. In N. Schmitt & F. Drasgow (Eds.), New Advances in Psychometric Methods. Jossey-Bass. San Francisco, CA.

Zuckerman, M., Joireman, J., Kraft, K. & Kuhlman, D.M. (1999). Where do motivational and emotional traits fit within three factor models of personality? Personality and Individual Differences, 26, 287-504.

Zumbo, B.D., Pope, G.A., Watson, J.E., & Hubley, A. M. (1997). An empirical test of Roskam's conjecture about the interpretation of an ICC parameter in personality inventories. Educational and Psychological Measurement, 57, 963-969.

Zwick, R., Donoghue, J.R., & Gima, A. (1993). Assessment of differential item functioning for performance tasks. Journal of Educational Measurement, 30, 233-251.

Zellars, K. L., Hochwarter, W. A., Perrewé, P. L. (2004). Experiencing job burnout: The roles of positive and negative traits and states. Journal of Applied Social Psychology, 34, 887-911

Appendix A

Table 1 Neuroticism Items from the IPIP-NEO

N1: ANXIETY
1. Worry about things.
7. Fear for the worst.
13. Am afraid of many things.
19. Get stressed out easily.
25. Get caught up in my problems.
31. Am not easily bothered by things.
37. Am relaxed most of the time.
43. Am not easily disturbed by events.
49. Don't worry about things that have already happened.
54. Adapt easily to new situations.
N2: ANGER
2. Get angry easily.
8. Get irritated easily.
14. Get upset easily.
20. Am often in a bad mood.
26. Lose my temper.
32. Rarely get irritated.
38. Seldom get mad.
44. Am not easily annoyed.
50. Keep my cool.
56. Rarely complain.
N3: DEPRESSION
3. Often feel blue.
9. Dislike myself.

15. Am often down in the dumps.
21. Have a low opinion of myself.
27. Have frequent mood swings.
33. Feel desperate.
39 .Feel that my life lacks direction.
45. Seldom feel blue.
51. Feel comfortable with myself.
57. Am very pleased with myself.
N4: SELF-CONSCIOUSNESS
4. Am easily intimidated.
10. Am afraid that I will do the wrong thing.
16. Find it difficult to approach others.
22. Am afraid to draw attention to myself.
28. Only feel comfortable with friends.
34. Stumble over my words.
40. Am not embarrassed easily.
46. Am comfortable in unfamiliar situations.
52. Am not bothered by difficult social situations.
58. Am able to stand up for myself.
N5: IMMODERATION
5. Often eat too much.
11. Don't know why I do some of the things I do.
17. Do things I later regret.
23. Go on binges.
29. Love to eat.
35. Rarely overindulge.
41. Easily resist temptations.

47. Am able to control my cravings.
53. Never spend more than I can afford.
59. Never splurge.
N6: VULNERABILITY
6. Panic easily.
12. Become overwhelmed by events.
18. Feel that I'm unable to deal with things.
24. Can't make up my mind.
30. Get overwhelmed by emotions.
36. Remain calm under pressure.
42. Can handle complex problems.
48. Know how to cope.
54. Readily overcome setbacks.
60. Am calm even in tense situations.

Appendix B

Table 2 Factor Analysis Results of the IPIP Neuroticism Scale

Component	Eigenvalue	Difference	Proportion of Variance	Cumulative Proportion
1	16.7577441	12.6305874	0.2793	0.2793
2	4.1271567	1.2899136	0.0688	0.3481
3	2.8372431	0.5325154	0.0473	0.3954
4	2.3047277	0.5344629	0.0384	0.4338
5	1.7702648	0.2071716	0.0295	0.4633
6	1.5630933	0.1205805	0.0261	0.4893
7	1.4425128	0.2276311	0.0240	0.5134
8	1.2148817	0.1507688	0.0202	0.5336
9	1.0641129	0.1528311	0.0177	0.5514
10	0.9112818	0.0496256	0.0152	0.5666
11	0.8616562	0.0140315	0.0144	0.5809
12	0.8476247	0.0350150	0.0141	0.5950
13	0.8126097	0.0270042	0.0135	0.6086
14	0.7856055	0.0406183	0.0131	0.6217
15	0.7449872	0.0181172	0.0124	0.6341
16	0.7268700	0.0172895	0.0121	0.6462
17	0.7095804	0.0050044	0.0118	0.6580
18	0.7045761	0.0275479	0.0117	0.6698
19	0.6770281	0.0081499	0.0113	0.6811
20	0.6688782		0.0111	0.6922

Figure 1 Scree plot for the Neuroticism Scale

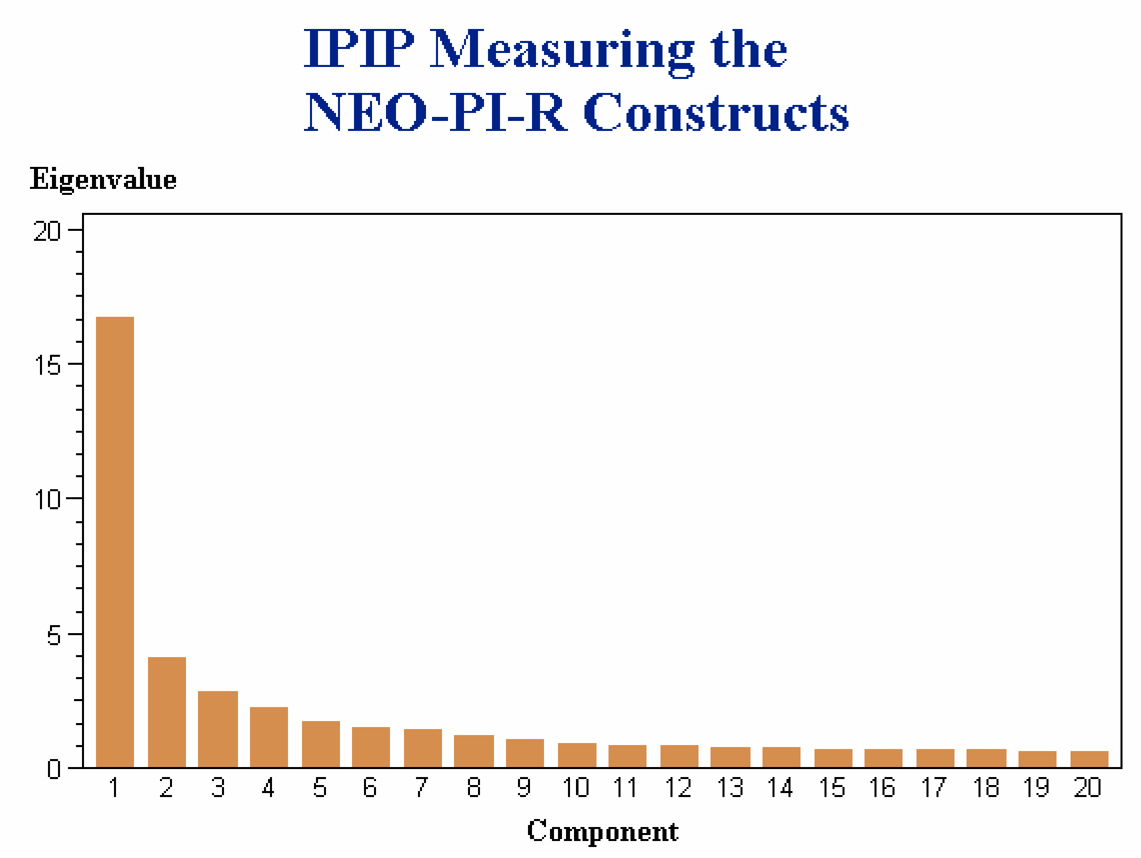


Table 3 Subscale Score Correlation Matrix

	Anxiety	Anger	Depression	Self-Consciousness	Immoderation	Vulnerability
Anxiety	1	.610(**) <.0001	.677(**) <.0001	.595(**) <.0001	.322(**) <.0001	.780(**) <.0001
Anger	.610(**) <.0001	1	.526(**) <.0001	.284(**) <.0001	.371(**) <.0001	.538(**) <.0001
Depression	.677(**) <.0001	.526(**) <.0001	1	.573(**) <.0001	.389(**) <.0001	.662(**) <.0001
Self-Consciousness	.595(**) <.0001	.284(**) <.0001	.573(**) <.0001	1	.202(**) <.0001	.599(**) <.0001
Immoderation	.322(**) <.0001	.371(**) <.0001	.389(**) <.0001	.202(**) <.0001	1	.366(**) <.0001
Vulnerability	.780(**) <.0001	.538(**) <.0001	.662(**) <.0001	.599(**) <.0001	.366(**) <.0001	1

n=20,993
** p<.001

Table 4 Item Correlation Matrix

	i1	i6	i11	i16	i21	i26	i31	i36	i41	i46
i1	1	0.263	0.362	0.308	0.096	0.458	0.461	0.278	0.297	0.372
		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i6	0.263	1	0.272	0.110	0.129	0.303	0.269	0.628	0.215	0.152
	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i11	0.362	0.272	1	0.283	0.121	0.331	0.417	0.315	0.510	0.333
	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
i16	0.308	0.110	0.283	1	0.128	0.416	0.320	0.162	0.336	0.398
	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001
i21	0.096	0.129	0.121	0.128	1	0.136	0.107	0.141	0.171	0.115
	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001
i26	0.458	0.303	0.331	0.416	0.136	1	0.472	0.332	0.332	0.371
	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
I31	0.461	0.269	0.417	0.320	0.107	0.472	1	0.328	0.402	0.433
	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001
I36	0.278	0.628	0.315	0.162	0.141	0.332	0.328	1	0.251	0.217
	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001
i41	0.297	0.215	0.510	0.336	0.171	0.332	0.402	0.251	1	0.389
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001
i46	0.372	0.152	0.333	0.398	0.115	0.371	0.433	0.217	0.389	1
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	
i51	0.160	0.211	0.262	0.190	0.137	0.233	0.222	0.237	0.289	0.284
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i56	0.338	0.214	0.312	0.371	0.118	0.470	0.371	0.246	0.300	0.368
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i61	0.417	0.223	0.367	0.447	0.145	0.536	0.480	0.267	0.397	0.423
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i66	0.375	0.610	0.376	0.268	0.146	0.462	0.383	0.585	0.316	0.281
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i71	0.358	0.280	0.731	0.299	0.139	0.368	0.452	0.332	0.580	0.370
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i76	0.182	0.064	0.291	0.360	0.073	0.225	0.266	0.151	0.320	0.317
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i81	0.194	0.271	0.291	0.207	0.152	0.255	0.265	0.283	0.323	0.290
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i86	0.320	0.219	0.440	0.353	0.112	0.437	0.399	0.260	0.467	0.384
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i91	0.505	0.401	0.435	0.348	0.139	0.566	0.465	0.448	0.375	0.388
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i96	0.288	0.438	0.529	0.211	0.125	0.325	0.400	0.463	0.454	0.261
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i101	0.299	0.192	0.494	0.374	0.165	0.347	0.406	0.234	0.734	0.418
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i1	i6	i11	i16	i21	i26	i31	i36	i41	i46
i106	0.192	0.017	0.234	0.347	0.057	0.218	0.245	0.070	0.320	0.310
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i111	0.100	0.160	0.202	0.095	0.317	0.155	0.145	0.180	0.203	0.117
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i116	0.234	0.118	0.270	0.321	0.108	0.319	0.274	0.174	0.296	0.341
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i121	0.435	0.266	0.451	0.314	0.131	0.414	0.427	0.325	0.378	0.388
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i126	0.241	0.720	0.275	0.106	0.149	0.312	0.284	0.598	0.233	0.167
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i131	0.326	0.432	0.525	0.233	0.143	0.380	0.374	0.462	0.405	0.290
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i136	0.172	0.102	0.260	0.260	0.062	0.192	0.261	0.158	0.263	0.260
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i141	0.052	0.068	0.033	0.065	0.556	0.062	0.036	0.089	0.032	0.049
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i146	0.383	0.288	0.404	0.314	0.131	0.463	0.367	0.286	0.329	0.336
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i151	0.403	0.326	0.325	0.252	0.086	0.400	0.355	0.345	0.273	0.272
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i156	0.280	0.577	0.298	0.134	0.114	0.300	0.288	0.609	0.234	0.174
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i161	0.300	0.233	0.516	0.302	0.126	0.367	0.404	0.275	0.510	0.345
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i166	0.138	0.084	0.190	0.307	0.099	0.216	0.195	0.127	0.242	0.256
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i171	0.078	0.129	0.127	0.082	0.388	0.119	0.077	0.140	0.140	0.066
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i176	0.303	0.261	0.240	0.314	0.087	0.489	0.310	0.258	0.262	0.266
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i181	0.387	0.284	0.350	0.207	0.056	0.371	0.354	0.304	0.332	0.251
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i186	0.244	0.612	0.259	0.095	0.114	0.280	0.256	0.509	0.216	0.133
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i191	0.175	0.140	0.390	0.244	0.109	0.215	0.278	0.196	0.414	0.278
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i196	0.266	0.109	0.214	0.379	0.075	0.299	0.252	0.143	0.256	0.314
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i201	0.086	0.142	0.143	0.129	0.218	0.132	0.095	0.154	0.153	0.101
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i206	0.149	0.111	0.128	0.263	0.071	0.267	0.180	0.106	0.196	0.189
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i1	i6	i11	i16	i21	i26	i31	i36	i41	i46
i211	0.318	0.193	0.229	0.267	0.083	0.370	0.272	0.199	0.209	0.246
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i216	0.260	0.537	0.277	0.136	0.109	0.300	0.283	0.573	0.232	0.170
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i221	0.302	0.238	0.590	0.227	0.088	0.289	0.361	0.270	0.437	0.286
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i226	0.185	0.094	0.188	0.262	0.054	0.215	0.214	0.128	0.208	0.215
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i231	0.134	0.152	0.183	0.153	0.382	0.184	0.149	0.183	0.222	0.138
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i236	0.273	0.213	0.347	0.278	0.084	0.376	0.329	0.228	0.384	0.271
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i241	0.429	0.197	0.317	0.270	0.074	0.374	0.381	0.223	0.303	0.316
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i246	0.271	0.466	0.279	0.195	0.123	0.381	0.290	0.402	0.272	0.200
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i251	0.293	0.189	0.445	0.321	0.139	0.314	0.363	0.224	0.644	0.360
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i256	0.227	0.092	0.237	0.312	0.080	0.251	0.240	0.141	0.242	0.255
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i261	0.059	0.088	0.082	0.036	0.134	0.072	0.029	0.093	0.079	0.016
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i266	0.203	0.143	0.265	0.255	0.076	0.272	0.267	0.163	0.300	0.220
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i271	0.245	0.143	0.257	0.302	0.070	0.325	0.300	0.193	0.290	0.265
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i276	0.240	0.354	0.229	0.135	0.133	0.258	0.246	0.380	0.194	0.150
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i281	0.267	0.158	0.442	0.295	0.110	0.276	0.341	0.193	0.633	0.341
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i286	0.193	0.002	0.230	0.441	0.068	0.286	0.242	0.061	0.320	0.280
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i291	0.058	0.077	0.035	0.018	0.164	0.059	-0.01	0.092	0.032	0.011
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i296	0.330	0.286	0.246	0.309	0.101	0.487	0.317	0.279	0.263	0.260
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i51	i56	i61	i66	i71	i76	i81	i86	i91	i96
i1	0.160	0.338	0.417	0.375	0.358	0.182	0.194	0.320	0.505	0.288
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i6	0.211	0.214	0.223	0.610	0.280	0.064	0.271	0.219	0.401	0.438
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i11	0.262	0.312	0.367	0.376	0.731	0.291	0.291	0.440	0.435	0.529
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i16	0.190	0.371	0.447	0.268	0.299	0.360	0.207	0.353	0.348	0.211
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i21	0.137	0.118	0.145	0.146	0.139	0.073	0.152	0.112	0.139	0.125
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i26	0.233	0.470	0.536	0.462	0.368	0.225	0.255	0.437	0.566	0.325
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i31	0.222	0.371	0.480	0.383	0.452	0.266	0.265	0.399	0.465	0.400
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i36	0.237	0.246	0.267	0.585	0.332	0.151	0.283	0.260	0.448	0.463
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i41	0.289	0.300	0.397	0.316	0.580	0.320	0.323	0.467	0.375	0.454
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i46	0.284	0.368	0.423	0.281	0.370	0.317	0.290	0.384	0.388	0.261
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i51	1	0.259	0.259	0.264	0.291	0.148	0.463	0.309	0.274	0.255
		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i56	0.259	1	0.448	0.357	0.350	0.252	0.248	0.452	0.479	0.286
	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i61	0.259	0.448	1	0.387	0.415	0.307	0.281	0.465	0.469	0.343
	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
i66	0.264	0.357	0.387	1	0.423	0.171	0.322	0.374	0.562	0.474
	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001
i71	0.291	0.350	0.415	0.423	1	0.326	0.334	0.514	0.486	0.606
	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001
i76	0.148	0.252	0.307	0.171	0.326	1	0.171	0.323	0.244	0.265
	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
i81	0.463	0.248	0.281	0.322	0.334	0.171	1	0.341	0.311	0.328
	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001
i86	0.309	0.452	0.465	0.374	0.514	0.323	0.341	1	0.483	0.423
	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001
i91	0.274	0.479	0.469	0.562	0.486	0.244	0.311	0.483	1	0.461
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		
i96	0.255	0.286	0.343	0.474	0.606	0.265	0.328	0.423	0.461	1
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	
i101	0.297	0.332	0.419	0.321	0.578	0.363	0.334	0.508	0.404	0.454
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i51	i56	i61	i66	i71	i76	i81	i86	i91	i96
i106	0.102	0.241	0.288	0.123	0.268	0.470	0.110	0.280	0.211	0.196
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i111	0.227	0.143	0.162	0.183	0.226	0.059	0.244	0.188	0.188	0.202
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i116	0.310	0.316	0.341	0.247	0.310	0.256	0.293	0.373	0.354	0.248
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i121	0.289	0.387	0.410	0.416	0.496	0.248	0.338	0.448	0.524	0.410
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i126	0.253	0.231	0.239	0.600	0.319	0.078	0.335	0.264	0.429	0.478
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i131	0.335	0.335	0.360	0.517	0.573	0.208	0.350	0.414	0.504	0.589
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i136	0.114	0.214	0.247	0.174	0.281	0.476	0.148	0.266	0.222	0.271
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i141	0.072	0.059	0.063	0.081	0.044	0.016	0.077	0.027	0.079	0.042
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i146	0.307	0.460	0.418	0.471	0.446	0.175	0.308	0.432	0.520	0.346
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i151	0.144	0.323	0.353	0.455	0.361	0.169	0.200	0.326	0.487	0.333
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i156	0.196	0.222	0.249	0.544	0.326	0.115	0.256	0.248	0.428	0.458
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i161	0.301	0.357	0.408	0.353	0.586	0.300	0.352	0.524	0.421	0.490
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i166	0.213	0.246	0.256	0.157	0.226	0.304	0.223	0.270	0.217	0.193
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i171	0.149	0.093	0.107	0.150	0.140	0.018	0.181	0.117	0.140	0.130
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i176	0.180	0.385	0.391	0.373	0.284	0.190	0.221	0.377	0.468	0.290
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i181	0.133	0.294	0.332	0.374	0.394	0.165	0.192	0.330	0.467	0.400
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i186	0.196	0.202	0.223	0.529	0.292	0.063	0.252	0.224	0.381	0.429
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i191	0.284	0.239	0.282	0.223	0.437	0.263	0.293	0.404	0.280	0.353
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i196	0.110	0.254	0.321	0.223	0.236	0.360	0.151	0.259	0.283	0.188
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i201	0.211	0.118	0.141	0.166	0.163	0.046	0.269	0.160	0.151	0.146
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i206	0.177	0.260	0.265	0.195	0.171	0.144	0.170	0.291	0.242	0.154
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i51	i56	i61	i66	i71	i76	i81	i86	i91	i96
i211	0.133	0.355	0.345	0.320	0.252	0.136	0.167	0.298	0.386	0.211
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i216	0.189	0.218	0.248	0.520	0.308	0.123	0.250	0.246	0.408	0.434
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i221	0.219	0.290	0.330	0.332	0.614	0.253	0.246	0.398	0.390	0.464
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i226	0.093	0.225	0.253	0.177	0.205	0.334	0.106	0.219	0.223	0.177
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i231	0.187	0.164	0.195	0.202	0.211	0.090	0.219	0.200	0.206	0.183
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i236	0.230	0.348	0.372	0.331	0.406	0.239	0.260	0.476	0.403	0.373
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i241	0.176	0.296	0.363	0.328	0.350	0.183	0.247	0.323	0.414	0.286
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i246	0.208	0.291	0.304	0.494	0.323	0.110	0.266	0.322	0.429	0.403
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i251	0.259	0.290	0.383	0.301	0.520	0.327	0.298	0.456	0.368	0.429
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i256	0.079	0.266	0.293	0.194	0.269	0.393	0.132	0.280	0.272	0.225
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i261	0.144	0.050	0.046	0.093	0.096	-.038	0.150	0.066	0.083	0.085
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i266	0.173	0.285	0.306	0.242	0.313	0.212	0.192	0.362	0.285	0.282
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i271	0.122	0.310	0.360	0.257	0.304	0.373	0.163	0.355	0.328	0.292
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i276	0.177	0.197	0.237	0.373	0.255	0.066	0.232	0.210	0.338	0.329
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i281	0.228	0.263	0.344	0.264	0.513	0.301	0.268	0.432	0.337	0.401
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i286	0.133	0.264	0.345	0.164	0.276	0.334	0.155	0.340	0.249	0.193
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i291	0.102	0.008	0.017	0.071	0.047	-.043	0.106	0.022	0.071	0.022
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i296	0.171	0.372	0.387	0.390	0.289	0.190	0.219	0.362	0.471	0.299
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i101	i106	i111	i116	i121	i126	i131	i136	i141	i146
i1	0.299	0.192	0.100	0.234	0.435	0.241	0.326	0.172	0.052	0.383
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
I6	0.192	0.017	0.160	0.118	0.266	0.720	0.432	0.102	0.068	0.288
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i11	0.494	0.234	0.202	0.270	0.451	0.275	0.525	0.260	0.033	0.404
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i16	0.374	0.347	0.095	0.321	0.314	0.106	0.233	0.260	0.065	0.314
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i21	0.165	0.057	0.317	0.108	0.131	0.149	0.143	0.062	0.556	0.131
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i26	0.347	0.218	0.155	0.319	0.414	0.312	0.380	0.192	0.062	0.463
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i31	0.406	0.245	0.145	0.274	0.427	0.284	0.374	0.261	0.036	0.367
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i36	0.234	0.070	0.180	0.174	0.325	0.598	0.462	0.158	0.089	0.286
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i41	0.734	0.320	0.203	0.296	0.378	0.233	0.405	0.263	0.032	0.329
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i46	0.418	0.310	0.117	0.341	0.388	0.167	0.290	0.260	0.049	0.336
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i51	0.297	0.102	0.227	0.310	0.289	0.253	0.335	0.114	0.072	0.307
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i56	0.332	0.241	0.143	0.316	0.387	0.231	0.335	0.214	0.059	0.460
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i61	0.419	0.288	0.162	0.341	0.410	0.239	0.360	0.247	0.063	0.418
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i66	0.321	0.123	0.183	0.247	0.416	0.600	0.517	0.174	0.081	0.471
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i71	0.578	0.268	0.226	0.310	0.496	0.319	0.573	0.281	0.044	0.446
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i76	0.363	0.470	0.059	0.256	0.248	0.078	0.208	0.476	0.016	0.175
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i81	0.334	0.110	0.244	0.293	0.338	0.335	0.350	0.148	0.077	0.308
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i86	0.508	0.280	0.188	0.373	0.448	0.264	0.414	0.266	0.027	0.432
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i91	0.404	0.211	0.188	0.354	0.524	0.429	0.504	0.222	0.079	0.520
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i96	0.454	0.196	0.202	0.248	0.410	0.478	0.589	0.271	0.042	0.346
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i101	1	0.378	0.203	0.339	0.405	0.227	0.406	0.296	0.033	0.356
		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i101	i106	i111	i116	i121	i126	i131	i136	i141	i146
i106	0.378	1	0.020	0.204	0.188	0.027	0.128	0.379	-.020	0.140
	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i111	0.203	0.020	1	0.173	0.221	0.212	0.283	0.071	0.235	0.220
	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
i116	0.339	0.204	0.173	1	0.352	0.160	0.312	0.188	0.081	0.331
	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001
i121	0.405	0.188	0.221	0.352	1	0.314	0.472	0.221	0.084	0.505
	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001
i126	0.227	0.027	0.212	0.160	0.314	1	0.489	0.120	0.095	0.342
	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
i131	0.406	0.128	0.283	0.312	0.472	0.489	1	0.215	0.081	0.503
	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001
i136	0.296	0.379	0.071	0.188	0.221	0.120	0.215	1	0.016	0.168
	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001
i141	0.033	-.020	0.235	0.081	0.084	0.095	0.081	0.016	1	0.094
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001
i146	0.356	0.140	0.220	0.331	0.505	0.342	0.503	0.168	0.094	1
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	
i151	0.288	0.144	0.104	0.206	0.397	0.334	0.373	0.156	0.046	0.401
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i156	0.227	0.045	0.166	0.150	0.320	0.600	0.459	0.133	0.073	0.300
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i161	0.528	0.224	0.241	0.340	0.449	0.275	0.464	0.255	0.045	0.428
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i166	0.284	0.259	0.110	0.291	0.228	0.106	0.208	0.237	0.048	0.197
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i171	0.123	-.028	0.417	0.126	0.162	0.165	0.196	0.027	0.318	0.170
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i176	0.292	0.167	0.102	0.277	0.320	0.285	0.305	0.165	0.038	0.367
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i181	0.327	0.139	0.121	0.164	0.361	0.307	0.374	0.161	-.011	0.314
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i186	0.202	0.018	0.156	0.116	0.275	0.652	0.422	0.099	0.061	0.301
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i191	0.429	0.215	0.206	0.340	0.330	0.183	0.338	0.219	0.046	0.262
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i196	0.292	0.397	0.037	0.214	0.234	0.110	0.184	0.292	0.025	0.226
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i201	0.157	-.008	0.292	0.185	0.185	0.168	0.204	0.045	0.183	0.185
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i206	0.237	0.156	0.041	0.249	0.170	0.134	0.174	0.104	0.029	0.228
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i101	**i106**	**i111**	**i116**	**i121**	**i126**	**i131**	**i136**	**i141**	**i146**
i211	0.231	0.136	0.095	0.202	0.324	0.208	0.269	0.111	0.053	0.383
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i216	0.218	0.054	0.149	0.156	0.300	0.546	0.427	0.143	0.065	0.293
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i221	0.447	0.202	0.170	0.232	0.402	0.257	0.476	0.217	0.025	0.372
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i226	0.240	0.284	0.029	0.159	0.188	0.099	0.159	0.323	0.015	0.159
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.033	<.001
i231	0.211	0.056	0.353	0.182	0.216	0.185	0.231	0.080	0.297	0.220
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i236	0.400	0.202	0.147	0.279	0.358	0.239	0.344	0.199	-.002	0.359
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.808	<.001
i241	0.313	0.162	0.101	0.250	0.392	0.214	0.309	0.149	0.029	0.366
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i246	0.278	0.073	0.161	0.199	0.306	0.518	0.401	0.124	0.052	0.353
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i251	0.656	0.318	0.178	0.298	0.360	0.212	0.374	0.263	0.014	0.319
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i256	0.279	0.338	0.058	0.195	0.258	0.102	0.198	0.335	0.038	0.205
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i261	0.063	-.051	0.213	0.063	0.105	0.121	0.141	-.029	0.099	0.128
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i266	0.316	0.181	0.092	0.245	0.275	0.157	0.257	0.174	0.013	0.257
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.058	<.001
i271	0.325	0.325	0.054	0.226	0.275	0.167	0.243	0.344	0.002	0.230
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.778	<.001
i276	0.186	-.002	0.142	0.178	0.293	0.374	0.347	0.090	0.099	0.259
	<.001	.747	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i281	0.657	0.323	0.146	0.282	0.348	0.184	0.346	0.241	-.006	0.286
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.356	<.001
i286	0.363	0.319	0.072	0.277	0.234	0.010	0.172	0.236	0.001	0.219
	<.001	<.001	<.001	<.001	<.001	0.167	<.001	<.001	0.857	<.001
i291	0.018	-.075	0.245	0.048	0.088	0.096	0.100	-.042	0.169	0.090
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i296	0.285	0.166	0.101	0.250	0.328	0.310	0.315	0.163	0.049	0.363
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i151	**i156**	**i161**	**i166**	**i171**	**i176**	**i181**	**i186**	**i191**	**i196**
i1	0.403	0.280	0.300	0.138	0.078	0.303	0.387	0.244	0.175	0.266
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i6	0.326	0.577	0.233	0.084	0.129	0.261	0.284	0.612	0.140	0.109
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i11	0.325	0.298	0.516	0.190	0.127	0.240	0.350	0.259	0.390	0.214
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i16	0.252	0.134	0.302	0.307	0.082	0.314	0.207	0.095	0.244	0.379
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i21	0.086	0.114	0.126	0.099	0.388	0.087	0.056	0.114	0.109	0.075
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i26	0.400	0.300	0.367	0.216	0.119	0.489	0.371	0.280	0.215	0.299
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i31	0.355	0.288	0.404	0.195	0.077	0.310	0.354	0.256	0.278	0.252
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i36	0.345	0.609	0.275	0.127	0.140	0.258	0.304	0.509	0.196	0.143
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i41	0.273	0.234	0.510	0.242	0.140	0.262	0.332	0.216	0.414	0.256
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i46	0.272	0.174	0.345	0.256	0.066	0.266	0.251	0.133	0.278	0.314
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i51	0.144	0.196	0.301	0.213	0.149	0.180	0.133	0.196	0.284	0.110
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i56	0.323	0.222	0.357	0.246	0.093	0.385	0.294	0.202	0.239	0.254
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i61	0.353	0.249	0.408	0.256	0.107	0.391	0.332	0.223	0.282	0.321
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i66	0.455	0.544	0.353	0.157	0.150	0.373	0.374	0.529	0.223	0.223
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i71	0.361	0.326	0.586	0.226	0.140	0.284	0.394	0.292	0.437	0.236
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i76	0.169	0.115	0.300	0.304	0.018	0.190	0.165	0.063	0.263	0.360
	<.001	<.001	<.001	<.001	.009	<.001	<.001	<.001	<.001	<.001
i81	0.200	0.256	0.352	0.223	0.181	0.221	0.192	0.252	0.293	0.151
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i86	0.326	0.248	0.524	0.270	0.117	0.377	0.330	0.224	0.404	0.259
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i91	0.487	0.428	0.421	0.217	0.140	0.468	0.467	0.381	0.280	0.283
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i96	0.333	0.458	0.490	0.193	0.130	0.290	0.400	0.429	0.353	0.188
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i101	0.288	0.227	0.528	0.284	0.123	0.292	0.327	0.202	0.429	0.292
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	i151	i156	i161	i166	i171	i176	i181	i186	i191	i196
i106	0.144	0.045	0.224	0.259	-.028	0.167	0.139	0.018	0.215	0.397
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.011	<.001	<.001

i111	0.104	0.166	0.241	0.110	0.417	0.102	0.121	0.156	0.206	0.037
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i116	0.206	0.150	0.340	0.291	0.126	0.277	0.164	0.116	0.340	0.214
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i121	0.397	0.320	0.449	0.228	0.162	0.320	0.361	0.275	0.330	0.234
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i126	0.334	0.600	0.275	0.106	0.165	0.285	0.307	0.652	0.183	0.110
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i131	0.373	0.459	0.464	0.208	0.196	0.305	0.374	0.422	0.338	0.184
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i136	0.156	0.133	0.255	0.237	0.027	0.165	0.161	0.099	0.219	0.292
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i141	0.046	0.073	0.045	0.048	0.318	0.038	-.011	0.061	0.046	0.025
	<.001	<.001	<.001	<.001	<.001	<.001	.096	<.001	<.001	<.001
i146	0.401	0.300	0.428	0.197	0.170	0.367	0.314	0.301	0.262	0.226
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i151	1	0.411	0.299	0.125	0.106	0.362	0.402	0.351	0.188	0.269
		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i156	0.411	1	0.268	0.099	0.167	0.292	0.356	0.580	0.188	0.154
	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i161	0.299	0.268	1	0.286	0.147	0.300	0.348	0.242	0.446	0.219
	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
i166	0.125	0.099	0.286	1	0.075	0.213	0.120	0.093	0.245	0.229
	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001
i171	0.106	0.167	0.147	0.075	1	0.101	0.094	0.165	0.145	0.058
	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001
i176	0.362	0.292	0.300	0.213	0.101	1	0.356	0.273	0.208	0.261
	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
i181	0.402	0.356	0.348	0.120	0.094	0.356	1	0.325	0.204	0.213
	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001
i186	0.351	0.580	0.242	0.093	0.165	0.273	0.325	1	0.152	0.122
	<.001	<.001	<.001	<.001	<.001	<.001	<.001			
i191	0.188	0.188	0.446	0.245	0.145	0.208	0.204	0.152	1	0.163
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001
i196	0.269	0.154	0.219	0.229	0.058	0.261	0.213	0.122	0.163	1
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	
i201	0.133	0.169	0.188	0.117	0.347	0.160	0.096	0.160	0.198	0.089
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i206	0.182	0.123	0.210	0.221	0.051	0.363	0.139	0.138	0.202	0.177
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i151	i156	i161	i166	i171	i176	i181	i186	i191	i196
i211	0.414	0.243	0.249	0.133	0.118	0.341	0.311	0.237	0.144	0.263
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i216	0.391	0.600	0.260	0.099	0.145	0.291	0.335	0.549	0.169	0.175
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i221	0.333	0.304	0.461	0.168	0.142	0.239	0.346	0.307	0.339	0.216
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i226	0.188	0.139	0.187	0.171	0.021	0.223	0.182	0.108	0.148	0.291
	<.001	<.001	<.001	<.001	.002	<.001	<.001	<.001	<.001	<.001
i231	0.183	0.186	0.228	0.129	0.417	0.195	0.157	0.175	0.212	0.127
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i236	0.332	0.252	0.423	0.221	0.116	0.428	0.362	0.242	0.318	0.227
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i241	0.388	0.264	0.314	0.147	0.091	0.306	0.371	0.233	0.197	0.283
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i246	0.376	0.445	0.303	0.148	0.164	0.463	0.401	0.484	0.207	0.178
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i251	0.304	0.243	0.493	0.254	0.133	0.305	0.381	0.213	0.398	0.292
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i256	0.262	0.168	0.236	0.230	0.054	0.265	0.234	0.116	0.197	0.344
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i261	0.082	0.106	0.101	0.020	0.252	0.056	0.078	0.110	0.131	0.007
	<.001	<.001	<.001	.004	<.001	<.001	<.001	<.001	<.001	<.001
i266	0.252	0.188	0.323	0.188	0.091	0.336	0.250	0.169	0.281	0.204
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i271	0.277	0.210	0.297	0.225	0.049	0.354	0.295	0.168	0.230	0.308
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i276	0.302	0.408	0.228	0.098	0.173	0.266	0.268	0.383	0.169	0.128
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i281	0.278	0.220	0.458	0.227	0.105	0.281	0.352	0.190	0.422	0.258
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i286	0.198	0.066	0.306	0.269	0.062	0.284	0.207	0.033	0.253	0.316
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i291	0.069	0.101	0.025	0.017	0.302	0.018	0.026	0.100	0.056	0.002
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.748
i296	0.387	0.311	0.292	0.196	0.110	0.660	0.392	0.297	0.197	0.275
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i201	i206	i211	i216	i221	i226	i231	i236	i241	i246
i1	0.086	0.149	0.318	0.260	0.302	0.185	0.134	0.273	0.429	0.271
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i6	0.142	0.111	0.193	0.537	0.238	0.094	0.152	0.213	0.197	0.466
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i11	0.143	0.128	0.229	0.277	0.590	0.188	0.183	0.347	0.317	0.279
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i16	0.129	0.263	0.267	0.136	0.227	0.262	0.153	0.278	0.270	0.195
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i21	0.218	0.071	0.083	0.109	0.088	0.054	0.382	0.084	0.074	0.123
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i26	0.132	0.267	0.370	0.300	0.289	0.215	0.184	0.376	0.374	0.381
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i31	0.095	0.180	0.272	0.283	0.361	0.214	0.149	0.329	0.381	0.290
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i36	0.154	0.106	0.199	0.573	0.270	0.128	0.183	0.228	0.223	0.402
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i41	0.153	0.196	0.209	0.232	0.437	0.208	0.222	0.384	0.303	0.272
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i46	0.101	0.189	0.246	0.170	0.286	0.215	0.138	0.271	0.316	0.200
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i51	0.211	0.177	0.133	0.189	0.219	0.093	0.187	0.230	0.176	0.208
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i56	0.118	0.260	0.355	0.218	0.290	0.225	0.164	0.348	0.296	0.291
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i61	0.141	0.265	0.345	0.248	0.330	0.253	0.195	0.372	0.363	0.304
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i66	0.166	0.195	0.320	0.520	0.332	0.177	0.202	0.331	0.328	0.494
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i71	0.163	0.171	0.252	0.308	0.614	0.205	0.211	0.406	0.350	0.323
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i76	0.046	0.144	0.136	0.123	0.253	0.334	0.090	0.239	0.183	0.110
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i81	0.269	0.170	0.167	0.250	0.246	0.106	0.219	0.260	0.247	0.266
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i86	0.160	0.291	0.298	0.246	0.398	0.219	0.200	0.476	0.323	0.322
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i91	0.151	0.242	0.386	0.408	0.390	0.223	0.206	0.403	0.414	0.429
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i96	0.146	0.154	0.211	0.434	0.464	0.177	0.183	0.373	0.286	0.403
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i101	0.157	0.237	0.231	0.218	0.447	0.240	0.211	0.400	0.313	0.278
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i201	i206	i211	i216	i221	i226	i231	i236	i241	i246
i106	-.008	0.156	0.136	0.054	0.202	0.284	0.056	0.202	0.162	0.073
	0.221	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i111	0.292	0.041	0.095	0.149	0.170	0.029	0.353	0.147	0.101	0.161
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i116	0.185	0.249	0.202	0.156	0.232	0.159	0.182	0.279	0.250	0.199
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i121	0.185	0.170	0.324	0.300	0.402	0.188	0.216	0.358	0.392	0.306
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i126	0.168	0.134	0.208	0.546	0.257	0.099	0.185	0.239	0.214	0.518
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i131	0.204	0.174	0.269	0.427	0.476	0.159	0.231	0.344	0.309	0.401
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i136	0.045	0.104	0.111	0.143	0.217	0.323	0.080	0.199	0.149	0.124
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i141	0.183	0.029	0.053	0.065	0.025	0.015	0.297	-.002	0.029	0.052
	<.001	<.001	<.001	<.001	<.001	0.033	<.001	0.808	<.001	<.001
i146	0.185	0.228	0.383	0.293	0.372	0.159	0.220	0.359	0.366	0.353
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i151	0.133	0.182	0.414	0.391	0.333	0.188	0.183	0.332	0.388	0.376
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i156	0.169	0.123	0.243	0.600	0.304	0.139	0.186	0.252	0.264	0.445
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i161	0.188	0.210	0.249	0.260	0.461	0.187	0.228	0.423	0.314	0.303
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i166	0.117	0.221	0.133	0.099	0.168	0.171	0.129	0.221	0.147	0.148
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i171	0.347	0.051	0.118	0.145	0.142	0.021	0.417	0.116	0.091	0.164
	<.001	<.001	<.001	<.001	<.001	0.002	<.001	<.001	<.001	<.001
i176	0.160	0.363	0.341	0.291	0.239	0.223	0.195	0.428	0.306	0.463
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i181	0.096	0.139	0.311	0.335	0.346	0.182	0.157	0.362	0.371	0.401
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i186	0.160	0.138	0.237	0.549	0.307	0.108	0.175	0.242	0.233	0.484
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i191	0.198	0.202	0.144	0.169	0.339	0.148	0.212	0.318	0.197	0.207
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i196	0.089	0.177	0.263	0.175	0.216	0.291	0.127	0.227	0.283	0.178
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i201	1	0.129	0.148	0.163	0.143	0.056	0.458	0.173	0.116	0.190
		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i206	0.129	1	0.199	0.132	0.135	0.186	0.146	0.303	0.166	0.228
	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i201	i206	i211	i216	i221	i226	i231	i236	i241	i246
i211	0.148	0.199	1	0.258	0.255	0.183	0.179	0.300	0.350	0.302
	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
i216	0.163	0.132	0.258	1	0.288	0.147	0.182	0.264	0.259	0.444
	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001
i221	0.143	0.135	0.255	0.288	1	0.187	0.179	0.346	0.309	0.274
	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001
i226	0.056	0.186	0.183	0.147	0.187	1	0.096	0.220	0.191	0.149
	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
i231	0.458	0.146	0.179	0.182	0.179	0.096	1	0.230	0.158	0.222
	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001
i236	0.173	0.303	0.300	0.264	0.346	0.220	0.230	1	0.323	0.388
	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001
i241	0.116	0.166	0.350	0.259	0.309	0.191	0.158	0.323	1	0.296
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001
i246	0.190	0.228	0.302	0.444	0.274	0.149	0.222	0.388	0.296	1
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	
i251	0.170	0.243	0.239	0.239	0.423	0.245	0.236	0.443	0.334	0.311
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i256	0.086	0.173	0.257	0.163	0.238	0.338	0.136	0.269	0.249	0.189
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i261	0.263	0.045	0.083	0.091	0.086	-.027	0.244	0.084	0.063	0.111
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i266	0.148	0.285	0.238	0.185	0.279	0.200	0.171	0.415	0.265	0.272
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i271	0.089	0.286	0.266	0.211	0.255	0.387	0.141	0.387	0.275	0.275
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i276	0.176	0.151	0.211	0.399	0.234	0.110	0.197	0.230	0.243	0.346
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i281	0.150	0.241	0.228	0.205	0.422	0.228	0.202	0.396	0.304	0.277
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i286	0.110	0.264	0.204	0.076	0.214	0.231	0.146	0.339	0.232	0.160
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i291	0.235	-.005	0.071	0.096	0.049	-.027	0.221	0.024	0.042	0.064
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i296	0.164	0.330	0.358	0.319	0.254	0.244	0.204	0.435	0.331	0.500
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i251	i256	i261	i266	i271	i276	i281	i286	i291	i296
i1	0.293	0.227	0.059	0.203	0.245	0.240	0.267	0.193	0.058	0.330
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i6	0.189	0.092	0.088	0.143	0.143	0.354	0.158	0.002	0.077	0.286
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i11	0.445	0.237	0.082	0.265	0.257	0.229	0.442	0.230	0.035	0.246
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i16	0.321	0.312	0.036	0.255	0.302	0.135	0.295	0.441	0.018	0.309
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i21	0.139	0.080	0.134	0.076	0.070	0.133	0.110	0.068	0.164	0.101
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i26	0.314	0.251	0.072	0.272	0.325	0.258	0.276	0.286	0.059	0.487
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i31	0.363	0.240	0.029	0.267	0.300	0.246	0.341	0.242	-.010	0.317
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.134	<.001
i36	0.224	0.141	0.093	0.163	0.193	0.380	0.193	0.061	0.092	0.279
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i41	0.644	0.242	0.079	0.300	0.290	0.194	0.633	0.320	0.032	0.263
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i46	0.360	0.255	0.016	0.220	0.265	0.150	0.341	0.280	0.011	0.260
	<.001	<.001	0.019	<.001	<.001	<.001	<.001	<.001	0.108	<.001
i51	0.259	0.079	0.144	0.173	0.122	0.177	0.228	0.133	0.102	0.171
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i56	0.290	0.266	0.050	0.285	0.310	0.197	0.263	0.264	0.008	0.372
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.247	<.001
i61	0.383	0.293	0.046	0.306	0.360	0.237	0.344	0.345	0.017	0.387
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.013	<.001
i66	0.301	0.194	0.093	0.242	0.257	0.373	0.264	0.164	0.071	0.390
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i71	0.520	0.269	0.096	0.313	0.304	0.255	0.513	0.276	0.047	0.289
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i76	0.327	0.393	-.038	0.212	0.373	0.066	0.301	0.334	-.043	0.190
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i81	0.298	0.132	0.150	0.192	0.163	0.232	0.268	0.155	0.106	0.219
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i86	0.456	0.280	0.066	0.362	0.355	0.210	0.432	0.340	0.022	0.362
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001
i91	0.368	0.272	0.083	0.285	0.328	0.338	0.337	0.249	0.071	0.471
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i96	0.429	0.225	0.085	0.282	0.292	0.329	0.401	0.193	0.022	0.299
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001
i101	0.656	0.279	0.063	0.316	0.325	0.186	0.657	0.363	0.018	0.285
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.011	<.001

	i251	i256	i261	i266	i271	i276	i281	i286	i291	i296
i106	0.318	0.338	-.051	0.181	0.325	-.002	0.323	0.319	-.075	0.166
	<.001	<.001	<.001	<.001	<.001	0.747	<.001	<.001	<.001	<.001
i111	0.178	0.058	0.213	0.092	0.054	0.142	0.146	0.072	0.245	0.101
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i116	0.298	0.195	0.063	0.245	0.226	0.178	0.282	0.277	0.048	0.250
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i121	0.360	0.258	0.105	0.275	0.275	0.293	0.348	0.234	0.088	0.328
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i126	0.212	0.102	0.121	0.157	0.167	0.374	0.184	0.010	0.096	0.310
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.167	<.001	<.001
i131	0.374	0.198	0.141	0.257	0.243	0.347	0.346	0.172	0.100	0.315
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i136	0.263	0.335	-.029	0.174	0.344	0.090	0.241	0.236	-.042	0.163
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i141	0.014	0.038	0.099	0.013	0.002	0.099	-.006	0.001	0.169	0.049
	0.049	<.001	<.001	0.058	0.778	<.001	0.356	0.857	<.001	<.001
i146	0.319	0.205	0.128	0.257	0.230	0.259	0.286	0.219	0.090	0.363
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i151	0.304	0.262	0.082	0.252	0.277	0.302	0.278	0.198	0.069	0.387
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i156	0.243	0.168	0.106	0.188	0.210	0.408	0.220	0.066	0.101	0.311
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i161	0.493	0.236	0.101	0.323	0.297	0.228	0.458	0.306	0.025	0.292
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i166	0.254	0.230	0.020	0.188	0.225	0.098	0.227	0.269	0.017	0.196
	<.001	<.001	0.004	<.001	<.001	<.001	<.001	<.001	0.014	<.001
i171	0.133	0.054	0.252	0.091	0.049	0.173	0.105	0.062	0.302	0.110
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i176	0.305	0.265	0.056	0.336	0.354	0.266	0.281	0.284	0.018	0.660
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.011	<.001
i181	0.381	0.234	0.078	0.250	0.295	0.268	0.352	0.207	0.026	0.392
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i186	0.213	0.116	0.110	0.169	0.168	0.383	0.190	0.033	0.100	0.297
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i191	0.398	0.197	0.131	0.281	0.230	0.169	0.422	0.253	0.056	0.197
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i196	0.292	0.344	0.007	0.204	0.308	0.128	0.258	0.316	0.002	0.275
	<.001	<.001	0.277	<.001	<.001	<.001	<.001	<.001	0.748	<.001
i201	0.170	0.086	0.263	0.148	0.089	0.176	0.150	0.110	0.235	0.164
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
i206	0.243	0.173	0.045	0.285	0.286	0.151	0.241	0.264	-.005	0.330
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

	i251	**i256**	**i261**	**i266**	**i271**	**i276**	**i281**	**i286**	**i291**	**i296**
i211	0.239	0.257	0.083	0.238	0.266	0.211	0.228	0.204	0.071	0.358
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i216	0.239	0.163	0.091	0.185	0.211	0.399	0.205	0.076	0.096	0.319
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i221	0.423	0.238	0.086	0.279	0.255	0.234	0.422	0.214	0.049	0.254
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i226	0.245	0.338	-.027	0.200	0.387	0.110	0.228	0.231	-.027	0.244
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i231	0.236	0.136	0.244	0.171	0.141	0.197	0.202	0.146	0.221	0.204
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i236	0.443	0.269	0.084	0.415	0.387	0.230	0.396	0.339	0.024	0.435
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001
i241	0.334	0.249	0.063	0.265	0.275	0.243	0.304	0.232	0.042	0.331
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i246	0.311	0.189	0.111	0.272	0.275	0.346	0.277	0.160	0.064	0.500
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i251	1	0.287	0.090	0.347	0.359	0.202	0.687	0.365	0.025	0.315
		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
i256	0.287	1	-.008	0.243	0.391	0.128	0.277	0.310	0.006	0.291
	<.001		0.240	<.001	<.001	<.001	<.001	<.001	0.369	<.001
i261	0.090	-.008	1	0.060	-.013	0.134	0.069	0.021	0.422	0.071
	<.001	0.240		<.001	0.062	<.001	<.001	0.002	<.001	<.001
i266	0.347	0.243	0.060	1	0.352	0.186	0.337	0.297	0.006	0.338
	<.001	<.001	<.001		<.001	<.001	<.001	<.001	0.354	<.001
i271	0.359	0.391	-.013	0.352	1	0.180	0.323	0.329	-.032	0.375
	<.001	<.001	0.062	<.001		<.001	<.001	<.001	<.001	<.001
i276	0.202	0.128	0.134	0.186	0.180	1	0.181	0.078	0.135	0.289
	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
i281	0.687	0.277	0.069	0.337	0.323	0.181	1	0.344	0.017	0.294
	<.001	<.001	<.001	<.001	<.001	<.001		<.001	0.016	<.001
i286	0.365	0.310	0.021	0.297	0.329	0.078	0.344	1	-.031	0.279
	<.001	<.001	0.002	<.001	<.001	<.001	<.001		<.001	<.001
i291	0.025	0.006	0.422	0.006	-.032	0.135	0.017	-.031	1	0.047
	<.001	0.369	<.001	0.354	<.001	<.001	0.016	<.001		<.001
i296	0.315	0.291	0.071	0.338	0.375	0.289	0.294	0.279	0.047	1
	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	

Appendix C

Table 5 Item Parameter Estimates for Females

	a	***b_1***	***b_2***	***b_3***	***b_4***
1	1.34	-3.38	-1.84	-1.12	0.852
2	1.17	-2.11	-0.581	0.361	2.00
3	1.87	-1.71	-0.581	0.157	1.39
4	1.06	-2.14	-0.496	0.541	2.31
5	0.481	-4.62	-1.92	-0.158	2.88
6	1.6	-1.37	-0.203	0.56	1.85
7	1.62	-1.73	-0.521	0.194	1.46
8	1.38	-2.55	-1.07	-0.268	1.37
9	1.7	-0.786	0.167	0.988	2.21
10	1.28	-2.85	-1.36	-0.514	1.04
11	0.953	-2.91	-1.45	-0.527	1.26
12	1.37	-2.13	-0.568	0.457	2.12
13	1.6	-1.34	-0.021	0.774	2.16
14	1.9	-1.74	-0.499	0.14	1.34
15	2.32	-1.1	-0.19	0.436	1.42
16	0.903	-2.23	-0.423	0.436	2.44
17	1.13	-2.43	-0.77	0.139	1.83
18	1.93	-0.925	0.166	0.814	2.01
19	2.21	-1.71	-0.632	-0.139	0.924
20	2.02	-0.79	0.362	1.09	2.19
21	1.83	-0.807	0.086	0.633	1.71
22	0.681	-2.81	-0.473	0.814	3.22
23	0.651	-2.52	-0.955	0.578	2.86
24	1.01	-2.49	-0.775	0.074	1.94
25	1.82	-2.45	-1.09	-0.415	1.2
26	1.3	-1.62	-0.302	0.32	1.77
27	1.98	-1.57	-0.566	-0.019	1.06
28	0.794	-3.1	-0.736	0.261	2.34
29	0.443	-6.88	-4.07	-1.91	1.44
30	1.65	-2.11	-0.895	-0.236	0.992
31	1.48	-2.17	-0.712	-0.135	1.37
32	1.4	-2.34	-0.867	-0.205	1.43
33	1.87	-0.512	0.322	0.982	2.17
34	0.726	-2.27	-0.024	1.24	3.95
35	0.48	-6.08	-2.17	0.209	4.13
36	1.31	-1.6	0.377	1.15	2.73
37	1.38	-2.14	0.034	0.771	2.49
38	1.21	-2.18	-0.431	0.241	2.07
39	1.13	-1.6	-0.276	0.407	1.94
40	0.863	-2.98	-0.912	-0.208	1.75
41	0.551	-5.2	-1.43	0.206	3.79

42	0.731	-1.86	1.44	2.84	5.13
43	1	-3.44	-0.924	0.161	2.63
44	1.3	-2.31	-0.542	0.022	1.77
45	1.66	-1.84	-0.559	0.034	1.44
46	0.719	-3.57	-0.548	0.512	2.88
47	0.708	-3.98	-0.504	0.632	3.29
48	1.57	-1.21	0.831	1.56	2.86
49	1.29	-2.45	-0.848	-0.41	1.39
50	1.44	-1.61	0.544	1.3	2.86
51	1.69	-1.07	0.525	1.1	2.29
52	0.87	-3.11	-0.809	-0.015	2.12
53	0.367	-5.04	-1.54	-0.266	4.11
54	1.13	-2.36	0.287	1.46	3.34
55	1.11	-1.71	0.752	1.57	3.34
56	1.05	-3.14	-1.04	-0.059	2.09
57	1.46	-1.92	-0.159	0.765	2.07
58	0.861	-1	1.53	2.31	4.21
59	0.446	-8.07	-4.11	-1.88	2.44
60	1.38	-1.93	0.014	0.774	2.31

Table 6 Item Parameter Estimates for Males

	a	b_1	b_2	b_3	b_4
1	1.333	-2.958	-1.55	-0.8	1.113
2	0.977	-2.203	-0.745	0.23	2.073
3	1.784	-1.978	-0.948	-0.156	1.134
4	1.117	-1.805	-0.307	0.767	2.603
5	0.319	-5.642	-2.111	0.848	5.491
6	1.597	-1.006	0.101	0.94	2.246
7	1.46	-1.938	-0.752	-0.015	1.358
8	1.225	-2.784	-1.301	-0.447	1.267
9	1.852	-0.938	-0.073	0.652	1.889
10	1.245	-2.713	-1.356	-0.552	1.011
11	0.854	-3.152	-1.713	-0.706	1.246
12	1.382	-1.835	-0.441	0.604	2.297
13	1.588	-1.138	0.182	1.052	2.481
14	1.568	-1.754	-0.526	0.178	1.522
15	2.244	-1.376	-0.506	0.141	1.164
16	1	-2.499	-0.976	-0.139	1.664
17	1.068	-2.958	-1.244	-0.318	1.511
18	1.911	-1.072	0.006	0.685	1.96
19	2.195	-1.509	-0.546	-0.013	1.011
20	1.862	-1.237	-0.072	0.701	1.909
21	1.989	-0.945	-0.136	0.39	1.389
22	0.813	-2.672	-0.727	0.463	2.695
23	0.537	-3.152	-1.397	0.777	3.45
24	1.088	-2.305	-0.704	0.148	1.971
25	1.764	-2.499	-1.146	-0.443	1.154
26	1.156	-1.672	-0.431	0.23	1.777
27	1.784	-1.662	-0.703	-0.065	1.052
28	0.854	-3.54	-1.397	-0.367	1.593
29	0.387	-8.07	-4.591	-1.958	2.052
30	1.529	-1.825	-0.636	0.109	1.399
31	1.401	-2.009	-0.545	0.06	1.603
32	1.235	-2.295	-0.906	-0.231	1.409
33	2.048	-0.936	-0.123	0.516	1.685
34	0.803	-2.346	-0.37	0.828	3.409
35	0.423	-6.57	-2.142	0.709	5.297
36	1.411	-1.291	0.538	1.277	2.756
37	1.382	-1.999	0.112	0.838	2.409
38	1.098	-2.009	-0.411	0.249	2.062
39	1.225	-1.927	-0.714	-0.033	1.358
40	0.931	-2.652	-0.814	-0.089	1.818
41	0.626	-4.591	-1.509	-0.096	3.175
42	0.752	-1.217	1.797	3.256	5.318
43	1.166	-2.529	-0.348	0.518	2.685

44	1.225	-2.193	-0.569	0.051	1.726
45	1.715	-1.958	-0.804	-0.208	1.113
46	0.804	-3.315	-0.636	0.497	2.767
47	0.735	-3.438	-0.453	0.777	3.511
48	1.637	-1.325	0.527	1.318	2.593
49	1.176	-2.468	-0.677	-0.2	1.593
50	1.431	-1.307	0.55	1.277	2.664
51	1.94	-1.179	0.185	0.712	1.828
52	1	-2.856	-0.866	-0.117	1.797
53	0.287	-5.978	-1.784	-0.166	5.369
54	1.205	-2.223	0.091	1.164	2.848
55	1.205	-1.682	0.499	1.369	3.236
56	0.809	-3.223	-0.943	0.133	2.562
57	1.607	-1.835	-0.327	0.478	1.756
58	0.99	-0.925	1.369	2.164	3.807
59	0.368	-9.223	-4.733	-1.201	3.756
60	1.392	-1.601	0.287	1.032	2.481

Note: Item parameters transformed to scale of Females. Transformed coefficients: slope (A)=1.0204 intercept (K)= -0.2947.

Table 7 Item Parameter Estimates for Age 20 and Under (*n* =7333)

	a	b_1	b_2	b_3	b_4
1	1.461	-2.682	-1.389	-0.637	1.132
2	1.146	-1.96	-0.618	0.401	1.94
3	1.861	-1.56	-0.537	0.227	1.369
4	1.114	-1.779	-0.343	0.707	2.34
5	0.456	-4.509	-1.769	0.483	3.624
6	1.661	-1.037	-0.02	0.775	1.95
7	1.503	-1.703	-0.521	0.243	1.455
8	1.367	-2.445	-1.094	-0.244	1.265
9	1.724	-0.734	0.143	0.939	2.045
10	1.241	-2.825	-1.436	-0.534	0.961
11	0.77	-4.176	-2.369	-1.151	1.075
12	1.241	-2.131	-0.606	0.623	2.302
13	1.598	-1.265	0.009	0.902	2.216
14	1.787	-1.589	-0.446	0.217	1.331
15	2.26	-1.018	-0.161	0.45	1.36
16	0.918	-2.311	-0.656	0.22	2.121
17	0.977	-2.968	-1.237	-0.228	1.55
18	1.956	-0.999	0.061	0.778	1.912
19	2.271	-1.389	-0.525	-0.024	0.909
20	1.892	-0.83	0.276	1.037	2.064
21	1.84	-0.79	0.052	0.608	1.598
22	0.771	-2.121	-0.207	0.932	2.996
23	0.541	-2.835	-1.104	1.208	3.757
24	0.958	-2.93	-1.132	-0.131	1.864
25	1.861	-2.283	-0.989	-0.255	1.236
26	1.304	-1.503	-0.307	0.339	1.626
27	1.861	-1.484	-0.573	0.022	1.008
28	0.801	-3.139	-0.961	0.103	1.997
29	0.427	-6.744	-3.776	-1.513	1.807
30	1.588	-1.902	-0.809	-0.073	1.056
31	1.609	-1.617	-0.402	0.152	1.417
32	1.451	-1.912	-0.697	-0.056	1.369
33	1.735	-0.758	0.153	0.882	2.026
34	0.711	-2.521	-0.286	1.132	3.785
35	0.43	-6.63	-2.321	1.103	5.402
36	1.419	-1.322	0.4	1.198	2.625
37	1.524	-1.541	0.308	1.065	2.577
38	1.199	-1.874	-0.398	0.311	1.959
39	1.02	-1.693	-0.304	0.545	2.13
40	1.049	-2.178	-0.65	-0.035	1.541
41	0.546	-4.842	-1.436	0.34	3.871
42	0.728	-1.731	1.265	2.73	5.003

43	1.062	-2.606	-0.553	0.535	2.644
44	1.335	-1.95	-0.501	0.087	1.645
45	1.619	-1.655	-0.502	0.12	1.408
46	0.795	-3.206	-0.525	0.604	2.844
47	0.719	-3.301	-0.233	1.046	3.7
48	1.503	-1.132	0.742	1.588	2.92
49	1.188	-2.388	-0.812	-0.289	1.522
50	1.535	-1.218	0.603	1.35	2.654
51	1.777	-0.932	0.431	1.018	2.102
52	1.001	-2.568	-0.536	0.263	2.102
53	0.328	-4.195	-0.492	1.217	6.002
54	1.135	-2.35	0.147	1.522	3.414
55	1.209	-1.551	0.629	1.503	3.196
56	0.97	-2.92	-1.065	-0.024	2.083
57	1.598	-1.455	-0.035	0.755	1.893
58	0.917	-0.792	1.493	2.302	4.08
59	0.466	-7.077	-3.519	-0.49	3.405
60	1.493	-1.484	0.13	0.906	2.283

Note: Item parameters transformed to scale of 20-to-30 age group. Transformed coefficients: slope (A)= 0.9511 intercept (K)= 0.0284

Table 8 Item Parameter Estimates for Age 20 to 30 Group (*n*=8080)

	a	b_1	b_2	b_3	b_4
1	1.49	-2.92	-1.55	-0.927	0.833
2	1.11	-1.94	-0.501	0.386	2.09
3	1.78	-1.82	-0.675	0.114	1.42
4	1.13	-1.85	-0.285	0.723	2.48
5	0.494	-3.97	-1.46	0.248	3.24
6	1.7	-1.04	0.016	0.703	1.96
7	1.56	-1.73	-0.538	0.164	1.51
8	1.3	-2.58	-1.07	-0.285	1.4
9	1.75	-0.705	0.234	1.01	2.26
10	1.3	-2.71	-1.26	-0.477	1.09
11	0.932	-2.77	-1.35	-0.454	1.41
12	1.48	-1.75	-0.314	0.581	2.21
13	1.67	-1.12	0.167	0.868	2.23
14	1.84	-1.57	-0.408	0.214	1.46
15	2.29	-1.15	-0.212	0.416	1.43
16	0.892	-2.39	-0.522	0.341	2.31
17	1.06	-2.54	-0.743	0.179	2.04
18	1.94	-0.833	0.262	0.883	2.14
19	2.24	-1.49	-0.488	-0.026	1.02
20	1.85	-0.918	0.313	1.05	2.28
21	1.88	-0.717	0.164	0.681	1.75
22	0.743	-2.64	-0.438	0.768	3.11
23	0.596	-2.95	-1.21	0.496	2.9
24	1.06	-2.34	-0.639	0.157	1.97
25	1.82	-2.4	-1	-0.379	1.25
26	1.26	-1.42	-0.172	0.41	1.91
27	1.96	-1.56	-0.555	-0.02	1.09
28	0.779	-3.32	-0.915	0.093	2.23
29	0.428	-7.4	-4.32	-1.97	1.62
30	1.66	-1.84	-0.694	-0.072	1.14
31	1.55	-1.95	-0.552	-0.03	1.45
32	1.38	-2.14	-0.774	-0.154	1.45
33	1.85	-0.564	0.303	0.933	2.19
34	0.704	-2.44	-0.031	1.23	4.16
35	0.505	-5.84	-2.02	0.162	4.06
36	1.44	-1.26	0.569	1.25	2.79
37	1.48	-1.94	0.135	0.79	2.41
38	1.21	-1.87	-0.25	0.34	2.17
39	1.14	-1.72	-0.371	0.249	1.71
40	0.93	-2.63	-0.71	-0.056	1.83
41	0.586	-4.99	-1.48	-0.04	3.4
42	0.84	-1.23	1.64	2.9	4.89

43	1.16	-2.71	-0.5	0.345	2.59
44	1.34	-2.08	-0.431	0.1	1.77
45	1.74	-1.82	-0.565	0.027	1.43
46	0.801	-3.12	-0.437	0.551	2.73
47	0.813	-3.35	-0.42	0.598	3.08
48	1.64	-1.13	0.824	1.53	2.81
49	1.34	-2.27	-0.641	-0.258	1.43
50	1.5	-1.3	0.662	1.37	2.85
51	1.81	-1.01	0.491	1.03	2.2
52	0.98	-2.72	-0.654	0.043	2.01
53	0.335	-6.08	-2.28	-0.974	3.85
54	1.2	-2.09	0.367	1.42	3.19
55	1.19	-1.46	0.845	1.61	3.38
56	1.01	-2.91	-0.884	0.041	2.24
57	1.62	-1.7	-0.082	0.762	2.04
58	0.946	-0.831	1.52	2.3	4.04
59	0.443	-8.13	-4.26	-1.88	2.43
60	1.51	-1.55	0.254	0.934	2.34

Table 9 Item Parameter Estimates for Over 30 to 55 (n=5226)

	a	b_1	b_2	b_3	b_4
1	1.329	-3.107	-1.676	-1.018	0.965
2	1.12	-2.106	-0.486	0.388	2.225
3	1.809	-1.776	-0.667	0.027	1.395
4	1.19	-1.796	-0.328	0.582	2.255
5	0.469	-4.467	-2.006	-0.4	2.806
6	1.699	-1.126	0.012	0.733	1.945
7	1.679	-1.636	-0.457	0.214	1.555
8	1.339	-2.506	-0.97	-0.175	1.645
9	1.779	-0.806	0.122	0.896	2.235
10	1.359	-2.376	-1.023	-0.235	1.325
11	1	-2.316	-0.938	-0.015	1.705
12	1.649	-1.646	-0.337	0.516	2.005
13	1.719	-0.97	0.272	1.025	2.376
14	1.859	-1.646	-0.371	0.267	1.595
15	2.209	-1.176	-0.287	0.387	1.505
16	0.88	-2.206	-0.459	0.462	2.646
17	1.23	-2.266	-0.653	0.248	2.025
18	1.959	-0.781	0.301	0.89	2.095
19	2.439	-1.506	-0.447	0.054	1.135
20	1.949	-0.872	0.306	1.095	2.335
21	1.979	-0.772	0.067	0.603	1.655
22	0.764	-3.057	-0.777	0.502	2.866
23	0.736	-2.186	-0.809	0.383	2.666
24	1.17	-1.706	-0.156	0.636	2.365
25	1.889	-2.256	-0.992	-0.344	1.245
26	1.319	-1.626	-0.299	0.334	1.965
27	2.009	-1.426	-0.418	0.177	1.365
28	0.734	-3.497	-0.87	0.208	2.676
29	0.398	-7.918	-4.667	-2.186	1.845
30	1.789	-1.696	-0.528	0.08	1.325
31	1.519	-2.186	-0.653	-0.08	1.495
32	1.349	-2.416	-0.834	-0.161	1.635
33	2.039	-0.432	0.306	0.935	2.155
34	0.83	-1.726	0.162	1.225	3.736
35	0.535	-5.077	-1.616	0.025	3.586
36	1.409	-1.356	0.566	1.235	2.636
37	1.449	-2.256	-0.088	0.581	2.195
38	1.27	-2.086	-0.336	0.295	2.135
39	1.28	-1.596	-0.411	0.202	1.725
40	0.81	-3.247	-0.909	-0.1	2.075
41	0.609	-4.777	-1.186	0.288	3.746
42	0.797	-1.276	1.895	3.126	5.137

43	1.28	-2.746	-0.568	0.203	2.305
44	1.29	-2.256	-0.391	0.191	2.045
45	1.709	-1.856	-0.606	-0.062	1.355
46	0.707	-3.567	-0.427	0.621	3.016
47	0.743	-3.817	-0.492	0.54	3.006
48	1.719	-1.146	0.872	1.515	2.706
49	1.389	-2.226	-0.586	-0.187	1.525
50	1.519	-1.476	0.554	1.225	2.826
51	1.699	-1.106	0.579	1.115	2.365
52	0.902	-3.137	-0.986	-0.257	1.855
53	0.398	-5.347	-1.876	-0.816	3.556
54	1.18	-2.076	0.438	1.315	3.006
55	1.13	-1.676	0.743	1.545	3.286
56	0.984	-3.007	-0.716	0.249	2.516
57	1.439	-2.186	-0.275	0.684	2.095
58	0.961	-0.915	1.545	2.195	3.876
59	0.434	-8.699	-4.267	-2.396	2.285
60	1.409	-1.806	0.193	0.835	2.325

Note: Item parameters transformed to scale of 20-30 year olds. Transformed coefficients: slope (A)= 1.0004 intercept (K)=-0.0955.

Appendix D

Table 10 IPIP Scale-level Descriptive Statistics for Males and Females

Scale	N items	Females (n= 13249)			Males (n=7743)			Effect Size*
		M	SD	Range	M	SD	Range	
Overall Neuroticism	60	180.956	36.503	76-290	166.222	36.820	60-300	0.40
Anxiety	10	31.797	7.606	10-50	27.924	7.754	10-50	0.50
Anger	10	30.358	8.914	9-50	27.797	9.239	9-50	0.28
Depression	10	28.074	9.333	8-50	26.674	9.543	9-50	0.15
Self-Consciousness	10	29.976	7.607	9-50	28.399	7.572	9-50	0.21
Immoderation	10	33.392	6.942	10-50	31.736	6.791	10-50	0.24
Vulnerability	10	27.361	7.368	9-50	23.694	7.223	8-50	0.50

* calculated as Cohen's *d*

Items are on a 1-5 scale. Missing data was treated as unscored and accounted for less than 0 .5% of response by item.

Figure 2 Mean Score by Gender

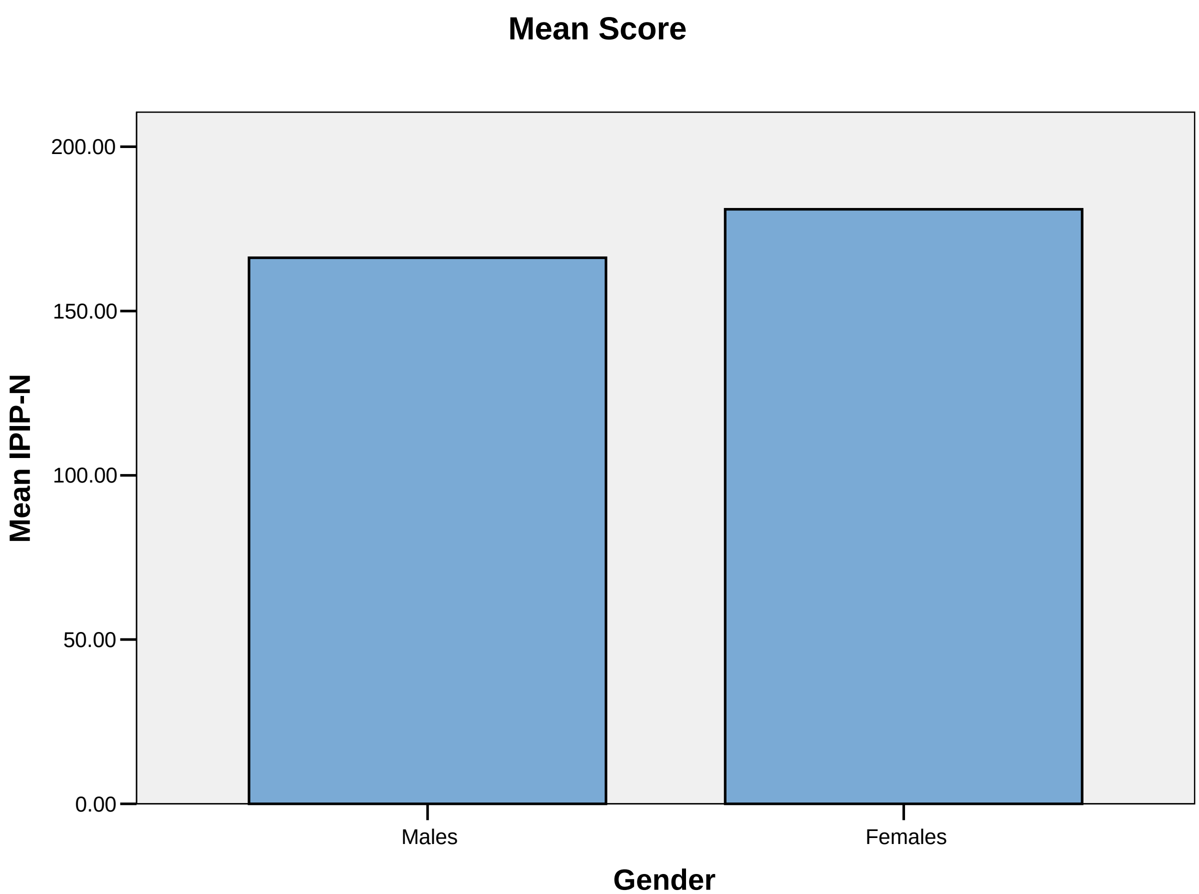


Table 11 IPIP Scale-level Descriptive Statistics and Comparison of Under-30 Age group and 20-30 year age group

Scale	N items	Under 30 (*n*=7333)			20-30 (*n*=6067)			Effect Size*
		M	SD	Range	M	SD	Range	
Overall Neuroticism	60	178.4526	35.608	60-290	177.2004	37.904	60-290	0.0340
Anxiety	10	30.6250	7.638	10-50	30.7631	7.997	10-50	-0.0177
Anger	10	30.1576	9.038	10-50	29.5819	9.252	9-50	0.0630
Depression	10	28.1143	9.374	9-50	27.8943	9.472	8-50	0.0233
Self-Consciousness	10	32.6233	6.520	10-50	33.3103	7.051	10-50	-0.1012
Immoderation	10	27.1058	7.172	9-50	26.0741	7.642	8-50	0.1392
Vulnerability	10	29.8265	7.543	9-50	29.5766	7.669	9-50	0.03285

* calculated as Cohen's *d*

Items are on a 1-5 scale. Missing data was treated as unscored and accounted for less than 0 .5% of response by item.

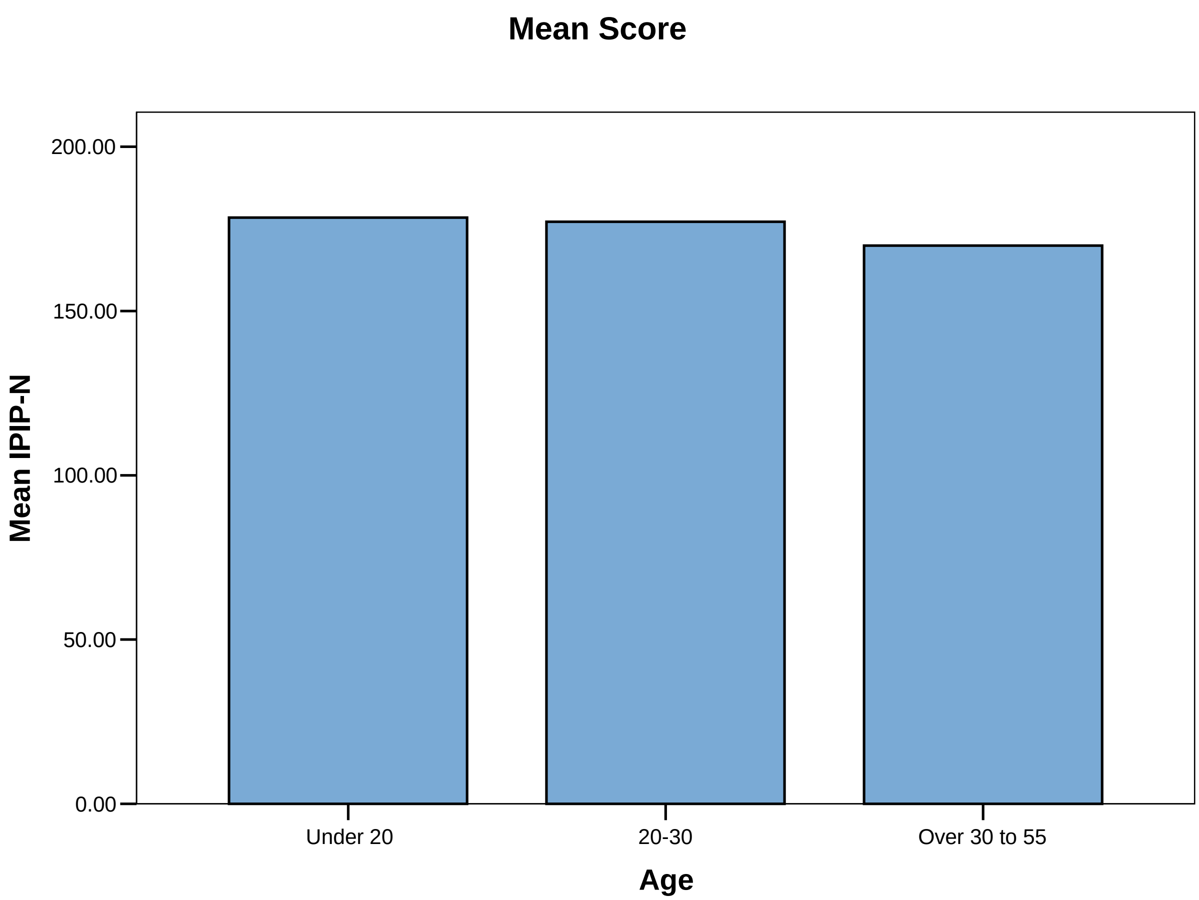
Table 12 IPIP Scale Descriptives and Comparison for 20-30 year-age group and over 30 age group

Scale	*n* items	20-30 (*n*=6067)			30-55 (*n*=5226)			Effect Size*
		M	SD	Range	M	SD	Range	
Overall Neuroticism	60	177.2004	37.904	60-290	169.9434	37.840	64-300	0.1916
Anxiety	10	30.7631	7.997	10-50	29.5683	7.977	10-50	0.1496
Anger	10	29.5819	9.252	9-50	28.3102	8.914	9-50	0.1400
Depression	10	27.8943	9.472	8-50	26.5522	9.346	950	0.1426
Self-Consciousness	10	33.3103	7.051	10-50	32.3703	7.225	11-50	0.1317
Immoderation	10	26.0741	7.642	8-50	24.5084	7.556	9-50	0.2060
Vulnerability	10	29.5766	7.669	9-50	28.6339	7.641	9-50	0.1232

* calculated as Cohen's *d*

Items are on a 1-5 scale. Missing data was treated as unscored and accounted for less than 0 .5% of response by item.

Figure 3 Mean Score by Age



Appendix E

Table 13 NCDIF results for Gender comparison

Item	**C-DIF**	**NC-DIF**	**CHI**	**PROB**
1	-0.305	0.032	65757.45	0
2	0.095	0.005	7659.59	0
3	0.489	0.079	74533.57	0
4	-0.273	0.024	482511.3	0
5	-0.429	0.061	31370.48	0
6	-0.541	0.09	227460.6	0
7	0.277	0.027	36078.74	0
8	0.171	0.012	20183.67	0
9	0.412	0.054	37389.77	0
10	0.005	0	3986.27	0
11	0.095	0.005	9512.23	0
12	-0.204	0.013	355565.2	0
13	-0.339	0.036	122129.9	0
14	-0.082	0.005	4437.08	0
15	0.578	0.105	82429.19	0
16	0.632	0.124	301559.4	0
17	0.479	0.076	112498.8	0
18	0.24	0.018	90057.97	0
19	-0.213	0.015	58249.98	0
20	0.692	0.148	149318.9	0
21	0.422	0.056	36956.57	0
22	0.237	0.019	26637.94	0
23	0.036	0.004	3522.27	0
24	-0.075	0.002	14920.82	0
25	0.045	0.001	31186.02	0
26	0.103	0.005	11802.8	0
27	0.098	0.006	7894.9	0
28	0.617	0.119	*********	0
29	-0.096	0.003	15829.33	0
30	-0.522	0.084	211373.5	0
31	-0.283	0.025	241565	0
32	-0.01	0	3084.1	0.1391
33	0.809	0.201	64966.23	0
34	0.277	0.024	73261.99	0
35	-0.196	0.012	55642.44	0
36	-0.231	0.017	136274.6	0
37	-0.082	0.002	51933.67	0
38	-0.047	0.001	40429.44	0
39	0.622	0.119	239165.6	0
40	-0.107	0.005	20446.29	0
41	0.155	0.008	27256.54	0

42	-0.304	0.029	*********	0
43	-0.443	0.068	46497.76	0
44	-0.016	0	25316.25	0
45	0.401	0.05	168485.6	0
46	0.025	0.001	3740	0
47	-0.125	0.005	210508.3	0
48	0.262	0.023	27900.43	0
49	-0.251	0.019	109384.6	0
50	-0.057	0.003	5801.24	0
51	0.448	0.069	18906.95	0
52	0.133	0.007	10864.84	0
53	-0.088	0.004	7046.63	0
54	0.184	0.012	15091.13	0
55	0.136	0.007	16500.2	0
56	-0.231	0.018	19753.09	0
57	0.273	0.025	20114.33	0
58	-0.012	0.001	3163.24	0.0188
59	-0.282	0.025	104019.8	0
60	-0.339	0.036	346744.3	0

********* = values greater than 999,999.

Table 14 NCDIF results for Younger/Mid comparison

Item	C-DIF	NC-DIF	CHI	PROB
1	-0.025	0.031	156049.1	0
2	0.009	0.002	49070.14	0
3	-0.006	0.012	12914	0
4	0.006	0.001	14985.15	0
5	-0.004	0.001	3012.47	0.4277
6	-0.001	0	3006.79	0.4566
7	-0.003	0.001	22632.92	0
8	0.008	0.001	4263.68	0
9	0.016	0.006	31342.4	0
10	0.008	0.005	56039.09	0
11	0.032	0.117	29501.23	0
12	0.005	0.017	9903.11	0
13	0.01	0.007	34707.98	0
14	0.007	0.001	60969.46	0
15	-0.002	0.003	8361.93	0
16	0.014	0.005	39022.35	0
17	0.039	0.073	255689	0
18	0.031	0.026	82462.38	0
19	0.012	0.002	4749.07	0
20	0.007	0.001	3058.96	0.2184
21	0.021	0.011	62108.03	0
22	-0.013	0.013	49015.58	0
23	-0.033	0.03	81415.7	0
24	0.024	0.048	45354.43	0
25	-0.003	0.002	13487.18	0
26	0.022	0.012	49628.28	0
27	0.001	0	3185.76	0.0089
28	0.007	0.001	7707.43	0
29	-0.014	0.011	169173.3	0
30	0.006	0.003	46088.59	0
31	-0.012	0.02	20668.89	0
32	0	0.004	6943.36	0
33	0.024	0.022	88686.33	0
34	0.014	0.006	209323.2	0
35	-0.027	0.017	30763.34	0
36	0.016	0.007	58771.66	0
37	-0.028	0.029	553754.9	0
38	0.013	0.005	105129.1	0
39	-0.031	0.019	19566.32	0
40	0.012	0.005	3021.92	0.3807
41	-0.017	0.009	155560	0
42	0.027	0.03	*********	0
43	-0.009	0.002	18946.26	0

44	0.008	0.001	9545.24	0
45	-0.009	0.004	60110.19	0
46	0	0	4949.14	0
47	-0.027	0.018	40084.59	0
48	0.005	0.001	16284.11	0
49	-0.003	0.003	4550.96	0
50	0.006	0.001	3317.68	0
51	0.007	0.001	4235.67	0
52	-0.013	0.008	170391.9	0
53	-0.078	0.246	*********	0
54	0.007	0.005	21675.56	0
55	0.017	0.007	45598.42	0
56	0.009	0.004	359302.5	0
57	0.002	0.002	4883.88	0
58	0.002	0	34757.2	0
59	-0.039	0.069	761946.9	0
60	0.011	0.002	14965.53	0

********* = values greater than 999,999.

Table 15 NCDIF results for Older/mid comparison

Item	C-DIF	NC-DIF	CHI	PROB
1	0.005	0.002	3165.09	0.0172
2	-0.001	0	3060.5	0.2127
3	-0.012	0.001	10862.24	0
4	-0.022	0.003	11177.46	0
5	-0.079	0.033	462746.1	0
6	-0.006	0	6488.94	0
7	0.025	0.004	20249.52	0
8	0.038	0.008	770048.4	0
9	-0.04	0.008	56525.04	0
10	0.074	0.03	82262.64	0
11	0.103	0.056	91454.2	0
12	-0.01	0.003	3886.54	0
13	0.051	0.013	187557.6	0
14	0.016	0.002	10532.46	0
15	-0.016	0.002	7652.59	0
16	0.036	0.007	252365.8	0
17	0.022	0.004	7446.37	0
18	0.013	0.001	18526.38	0
19	0.028	0.004	53356.36	0
20	0.013	0.001	95998.38	0
21	-0.033	0.006	20861.17	0
22	-0.066	0.022	*********	0
23	0.035	0.011	6683.02	0
24	0.147	0.113	193635.1	0
25	0.006	0.001	5607.77	0
26	-0.035	0.006	81157.16	0
27	0.081	0.034	124862.7	0
28	0.027	0.004	16867.4	0
29	0	0	3012.99	0.4251
30	0.064	0.021	83319.77	0
31	-0.025	0.004	11771.55	0
32	-0.004	0.002	3177.73	0.0115
33	0.028	0.005	9689.28	0
34	0.052	0.017	15058.1	0
35	0.007	0.001	3950.7	0
36	-0.014	0.001	133685.9	0
37	-0.072	0.027	346580.4	0
38	-0.026	0.004	405505.3	0
39	-0.002	0.001	3028.78	0.3476
40	-0.02	0.006	4633.15	0
41	0.047	0.012	*********	0
42	0.012	0.001	7621.37	0
43	-0.033	0.006	23812.28	0

44	0.022	0.004	7797.24	0
45	-0.025	0.003	126329.3	0
46	0	0.002	3002.96	0.4762
47	-0.024	0.003	23178.51	0
48	0.005	0	6560.07	0
49	0.021	0.002	122647.5	0
50	-0.036	0.007	246381.2	0
51	0.007	0.002	3447.18	0
52	-0.072	0.028	65397.04	0
53	0.008	0.001	4229.74	0
54	-0.003	0	4475.51	0
55	-0.034	0.006	397055.9	0
56	0.044	0.01	91483.62	0
57	-0.06	0.021	20080.37	0
58	-0.009	0	66128.66	0
59	-0.026	0.003	196983.7	0
60	-0.038	0.008	53712.91	0

********* = values greater than 999,999.

Appendix F

Boundary Response Functions (BRFs) of Items Demonstrating Gender DIF

Figure 4 Boundary response functions for Item 15 “Am often down in the dumps” from the Depression subscale. The solid lines represent the female group. The dotted lines represent the male group.

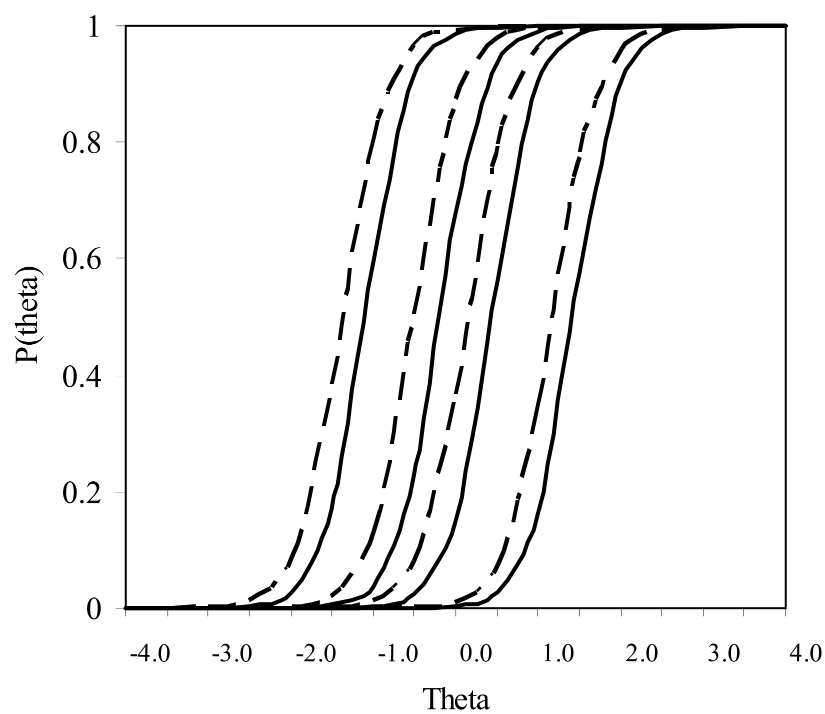


Figure 5 Boundary response functions for Item 16, "Find it difficult to approach others" from the self-Consciousness subscale. The solid lines represent the female group. The dotted lines represent the male group.

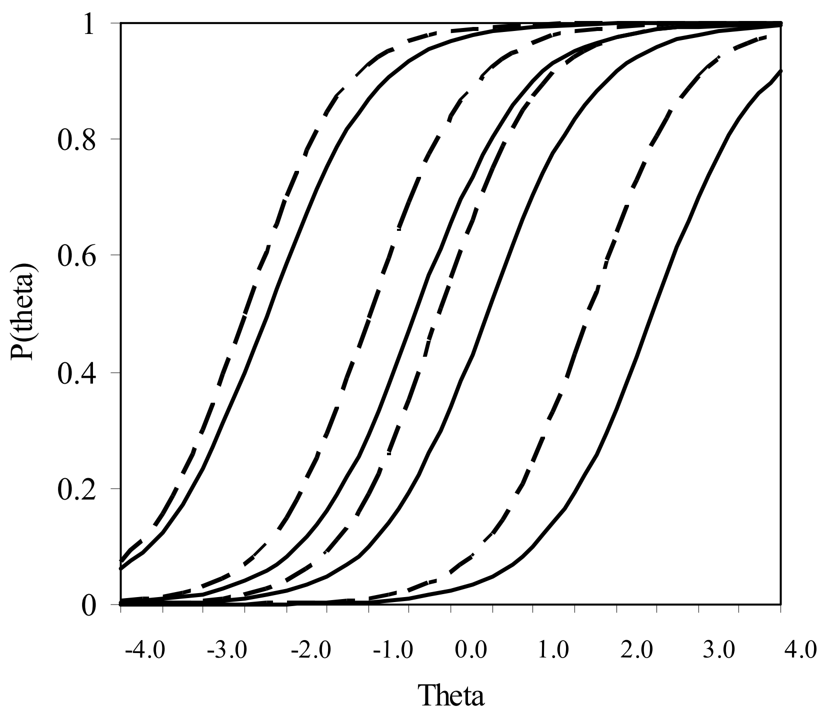


Figure 6 Boundary response functions for the item 20, “Am often in a bad mood” from the Anger subscale. The solid lines represent the female group. The dotted lines represent the male group.

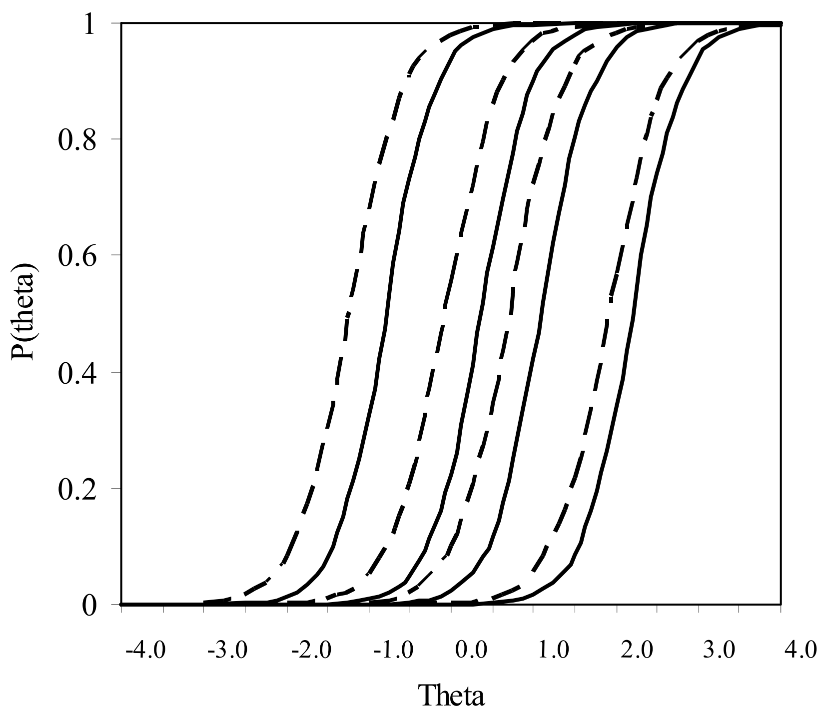


Figure 7 Boundary response functions for Item 28, “Only feel comfortable with friends” from the Self-Consciousness subscale. The solid lines represent the female group. The dotted lines represent the male group.

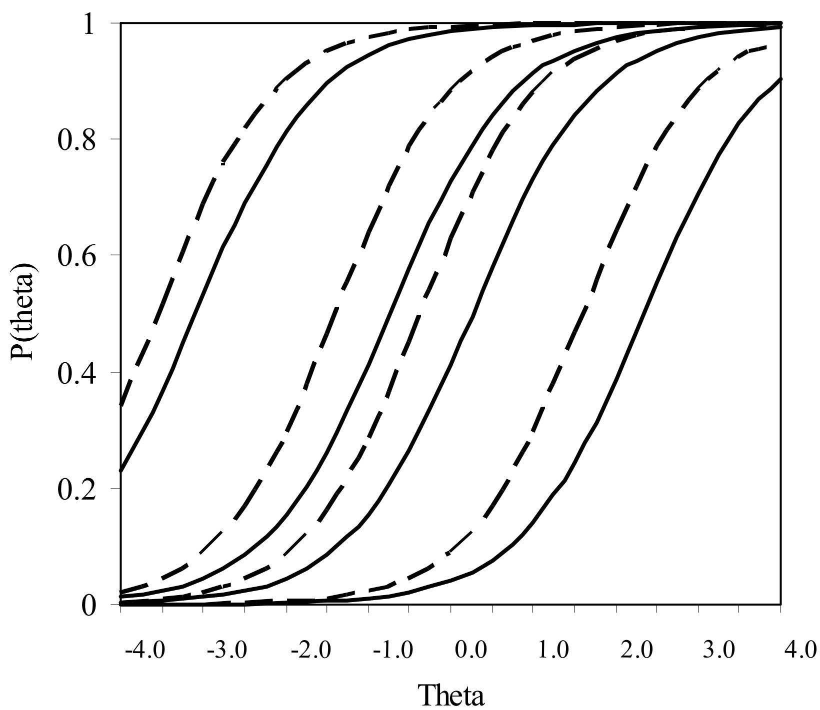


Figure 8 Boundary response functions for Item 33, “Feel desperate” from the Depression subscale. The solid lines represent the female group. The dotted lines represent the male group**.**

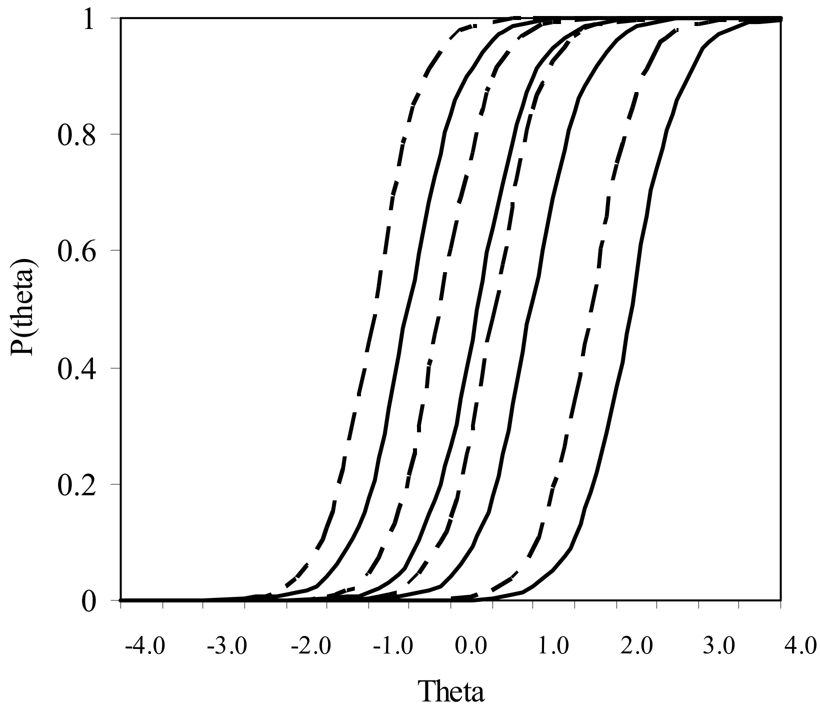
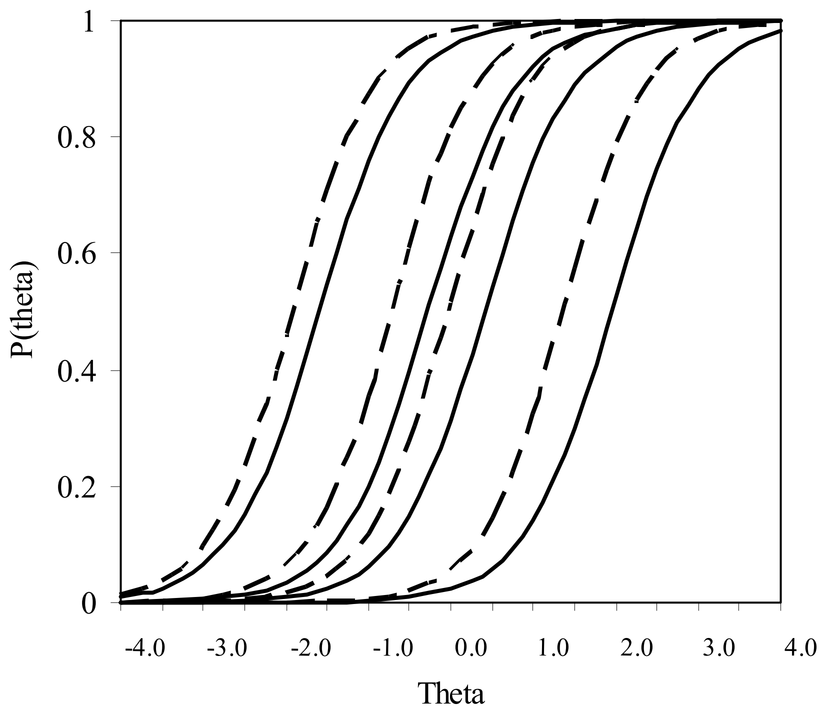


Figure 9 Boundary response functions for Item 39, “Feel that my life lacks direction” from the Depression subscale. The solid lines represent the female group. The dotted lines represent the male group.



Appendix G

Boundary Response Functions (BRFs) of Items Demonstrating Age DIF

Figure 10 Boundary response functions for Item 11, "Don't know why I do some of the things I do." Solid lines indicated the 20 to 30 age group, dotted lines indicated the younger the age of 20 group.

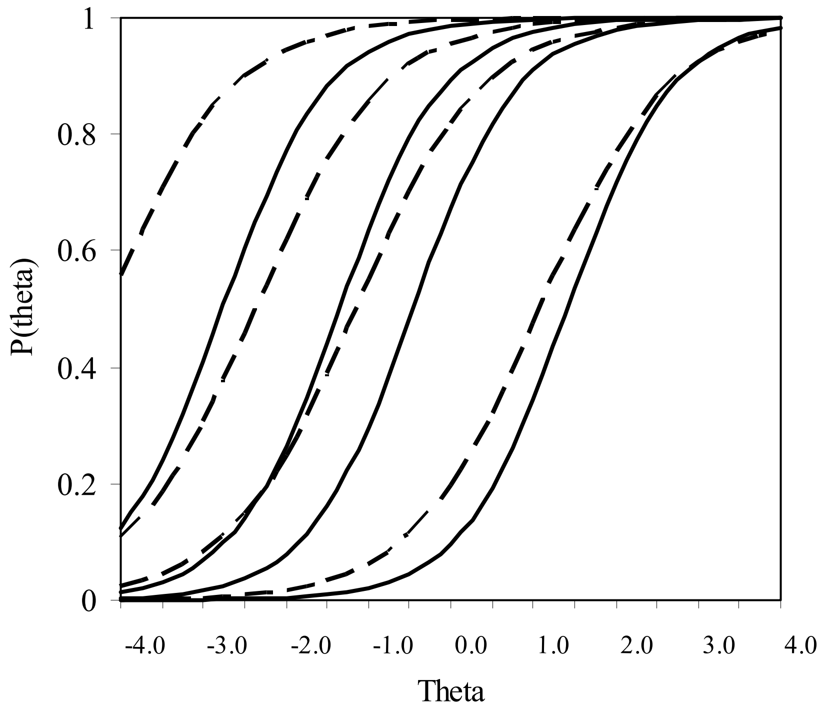


Figure 11 Boundary response functions for Item 53, “Never spend more than I can afford” from the Immoderation subscale. Solid lines indicated the age 20-to-30 group, dotted lines indicated the under 20 group.

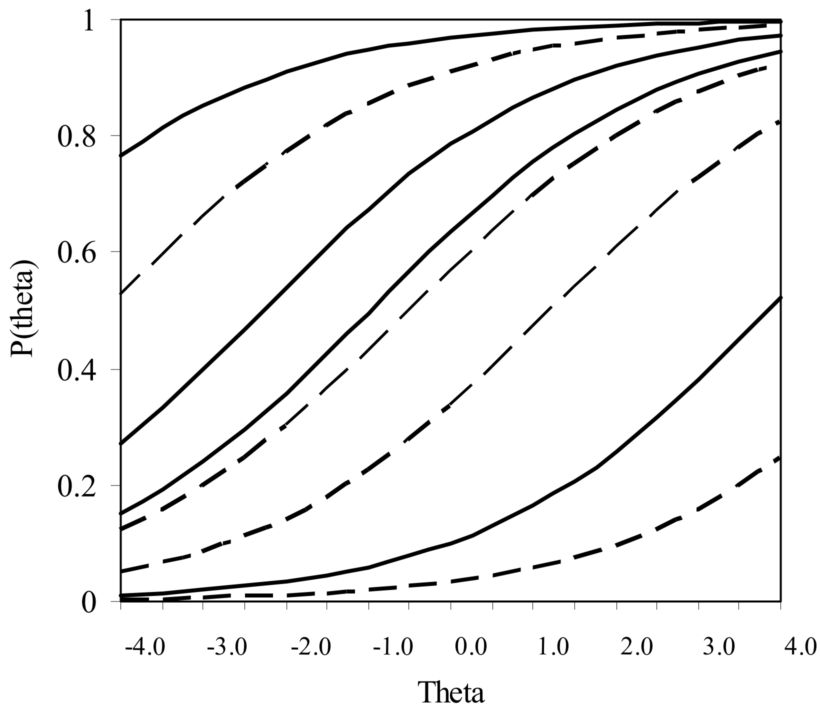
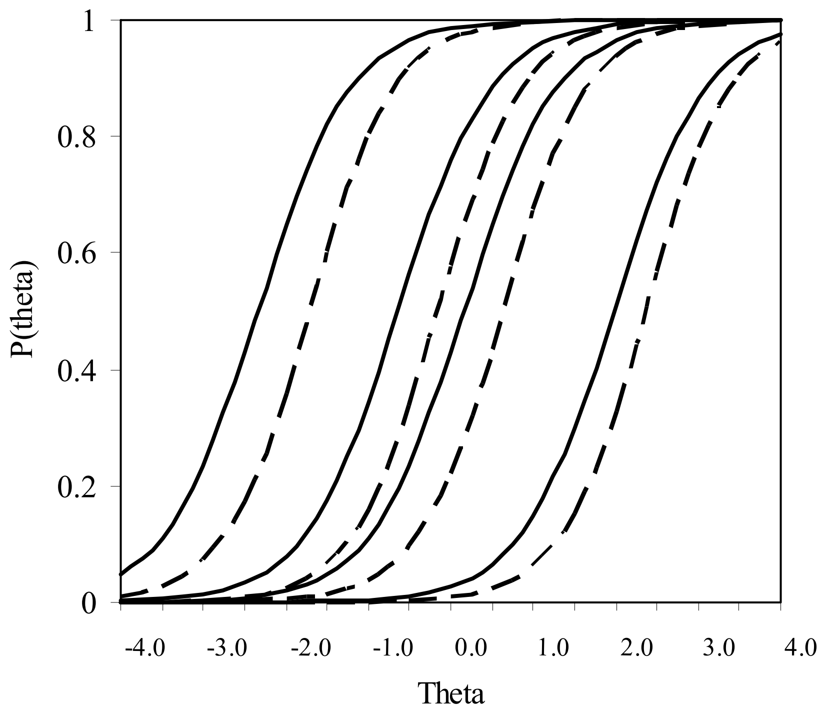


Figure 12 Boundary response functions for Item 24, "Readily overcome setbacks" from the Vulnerability subscale. Solid lines indicated the 20-to-30 age group, dotted lines indicated the over 30 to 55 age group.



Appendix H

Category Response Functions (CRFs) of Items Demonstrating Gender DIF

Figure 13 Category response functions for Item 15, “Am often down in the dumps” from the Depression subscale. The solid lines represent the female group. The dotted lines represent the male group.

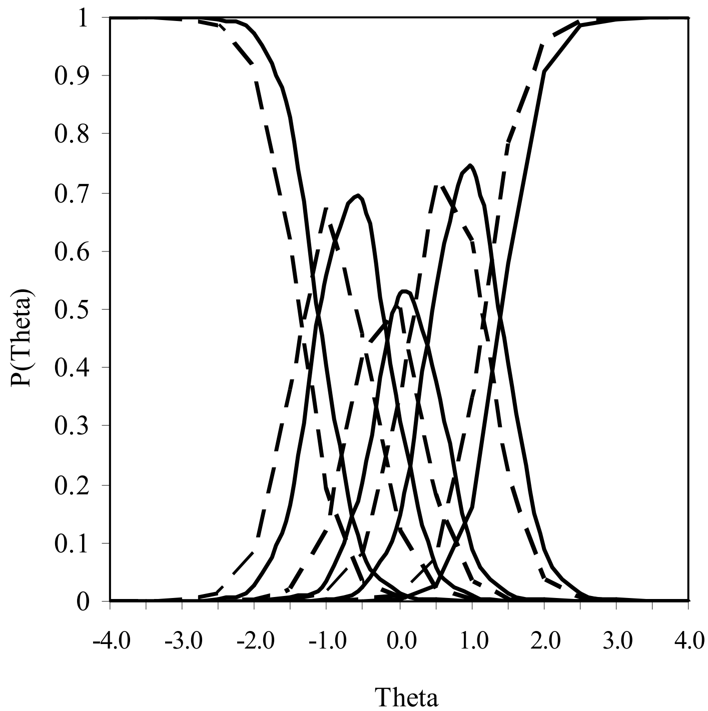


Figure 14 Category response functions for Item 16, “Find it difficult to approach others” From the Self-Consciousness subscale. The solid lines represent the female group. The dotted lines represent the male group.

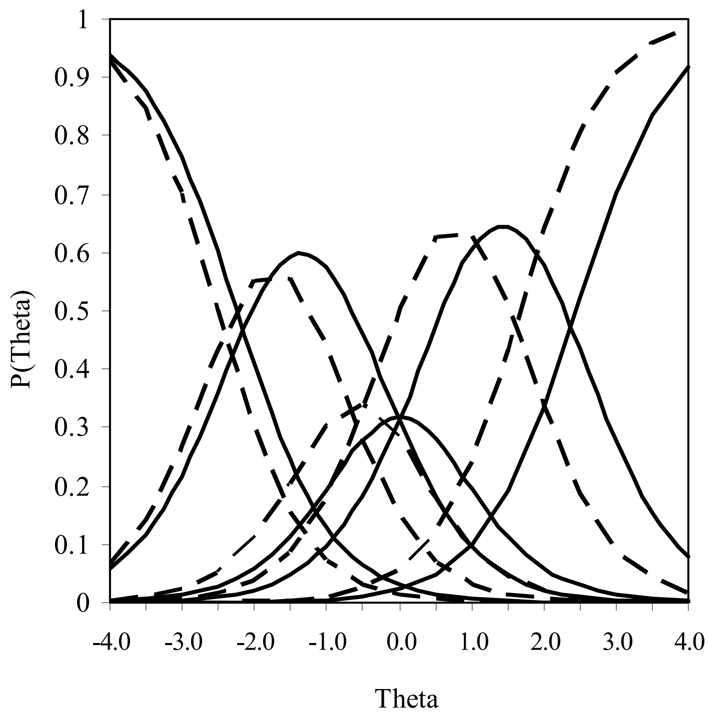


Figure 15 Category response functions for Item 20, “Am often in a bad mood” from the Anger subscale. The solid lines represent the female group. The dotted lines represent the male group.

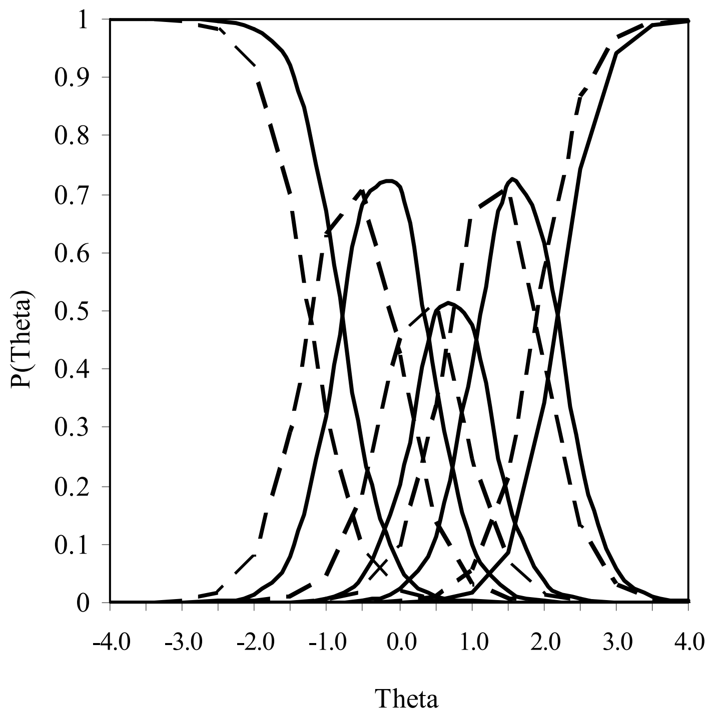


Figure 16 Category response functions for Item 28, "Only feel comfortable with friends" from the Self-Consciousness subscale. The solid lines represent the female group. The dotted lines represent the male group.

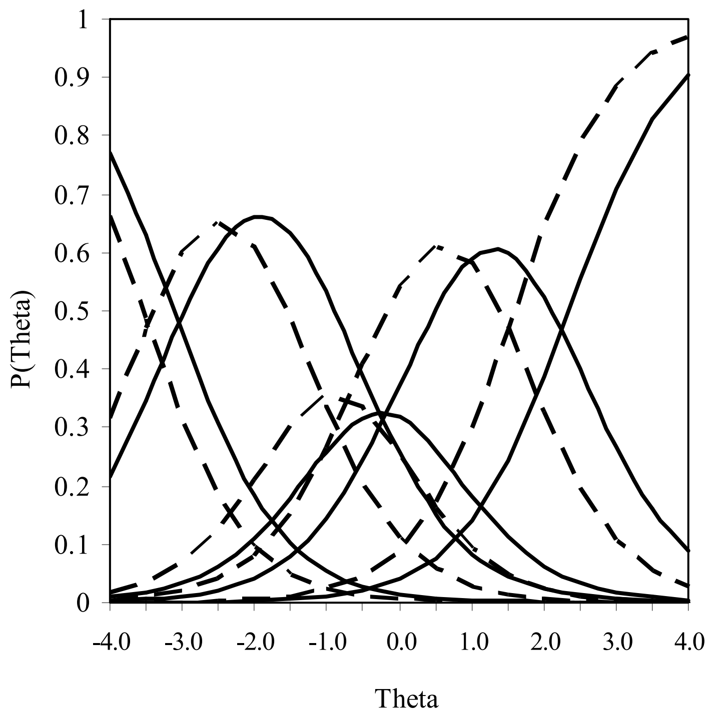


Figure 17 Category response functions for Item 33, “Feel desperate” from the Depression subscale. The solid lines represent the female group. The dotted lines represent the male group.

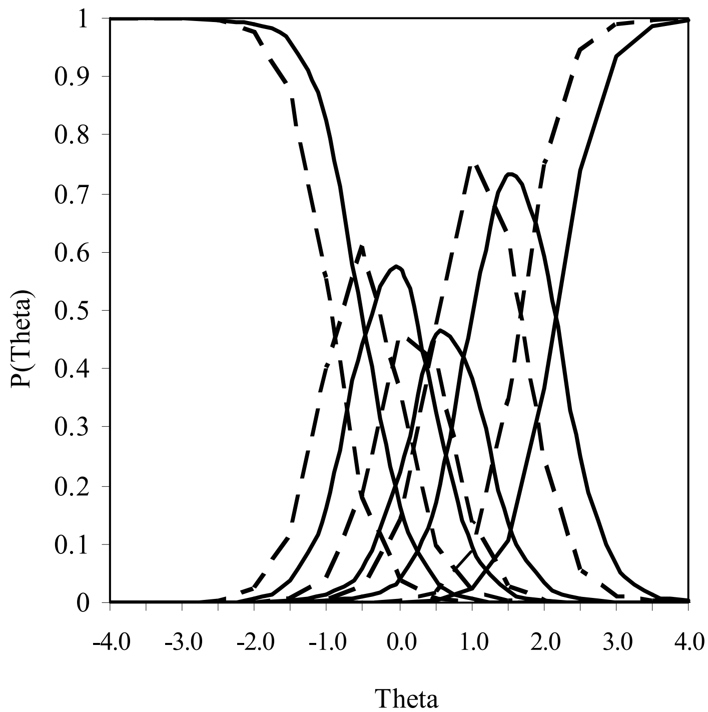
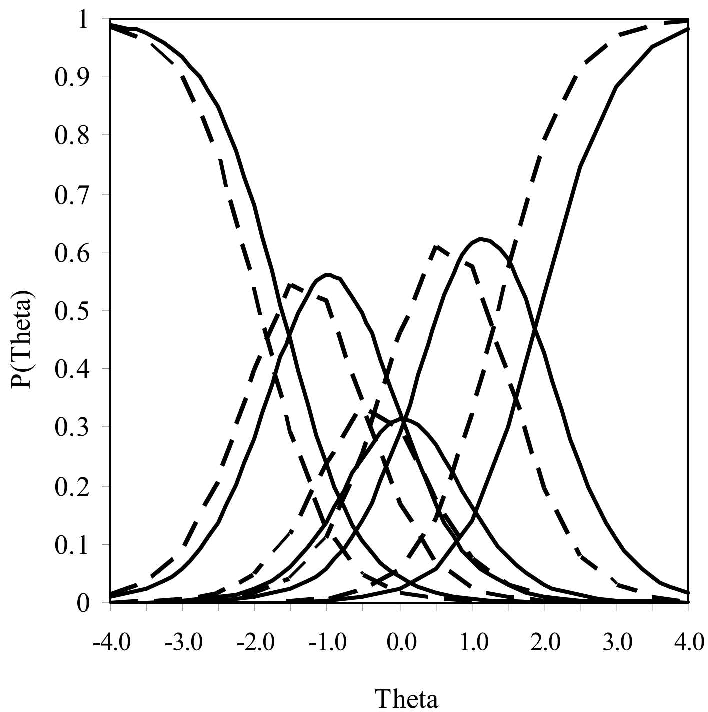


Figure 18 Category response functions for Item 39, "Feel that my life lacks direction" From the Depression subscale. The solid lines represent the female group. The dotted lines represent the male group.



Appendix I

Category Response Functions (CRFs) of Items Demonstrating Age DIF

Figure 19 Category response functions for Item 11, “Don't know why I do some of the things I do” From the Immoderation subscale. The solid lines represent the age 20 to age 30 group. The dotted lines represent the under 20 group.

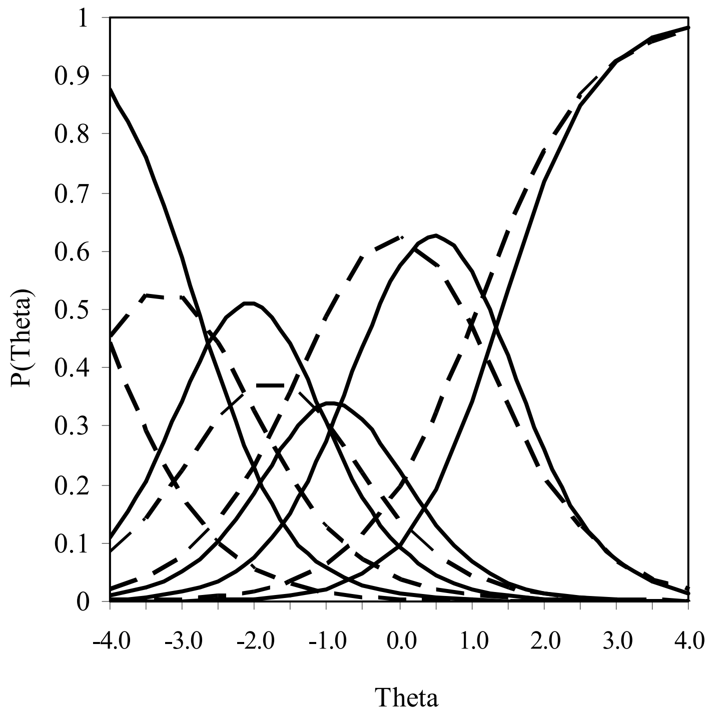


Figure 20 Category response functions for Item 53, “Never spend more than I can afford” from the Immoderation subscale. Solid lines represent the age 20-to-30 group. Dotted lines represent the under 20 group.

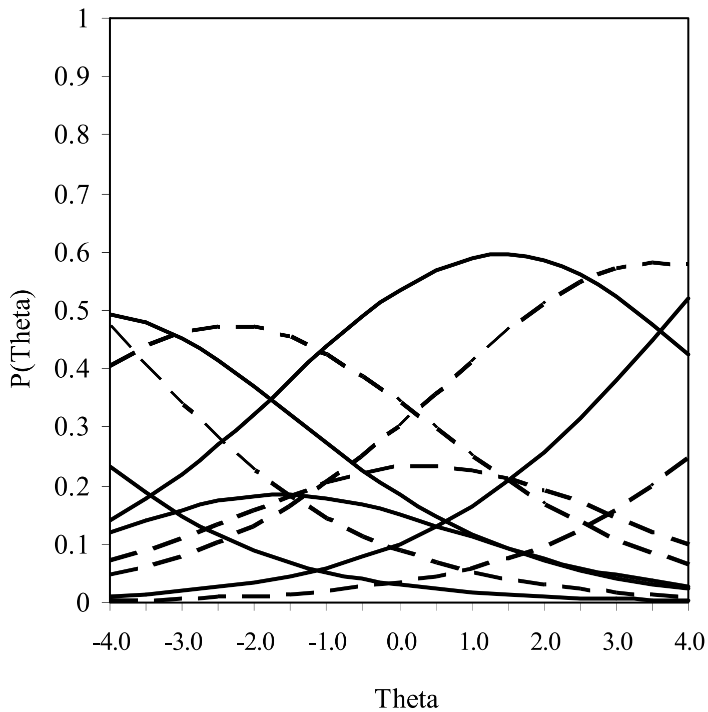
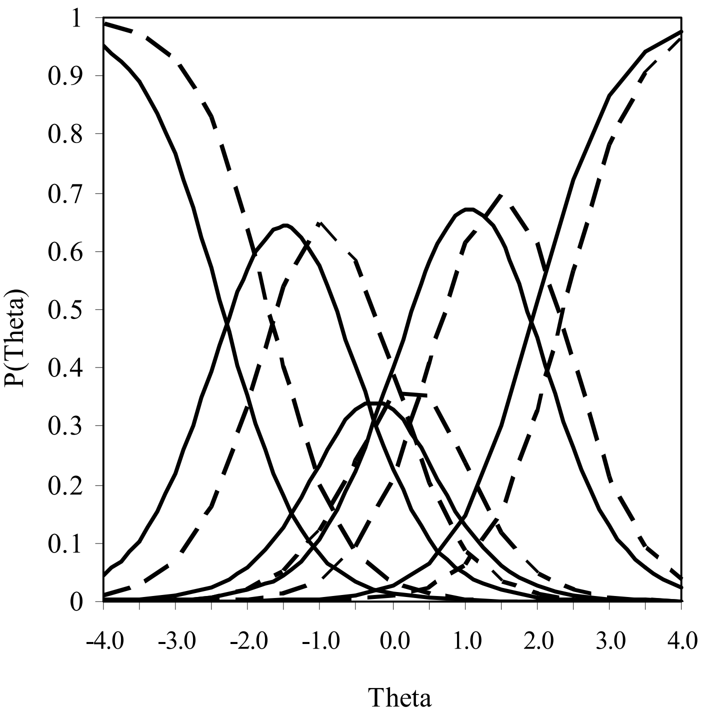


Figure 21 Category response functions for Item 24, Can't make up my mind from the Vulnerability subscale. The solid lines represent the age 20-to-30 age group. The dotted lines represent the over 30 to 55 age group.



Appendix J

Expected True Score Functions of Select Items

Figure 22 Expected Item Score for Item 33, "Feel desperate" from the Depression subscale. The solid lines represent the female group. The dotted lines represent the male group.

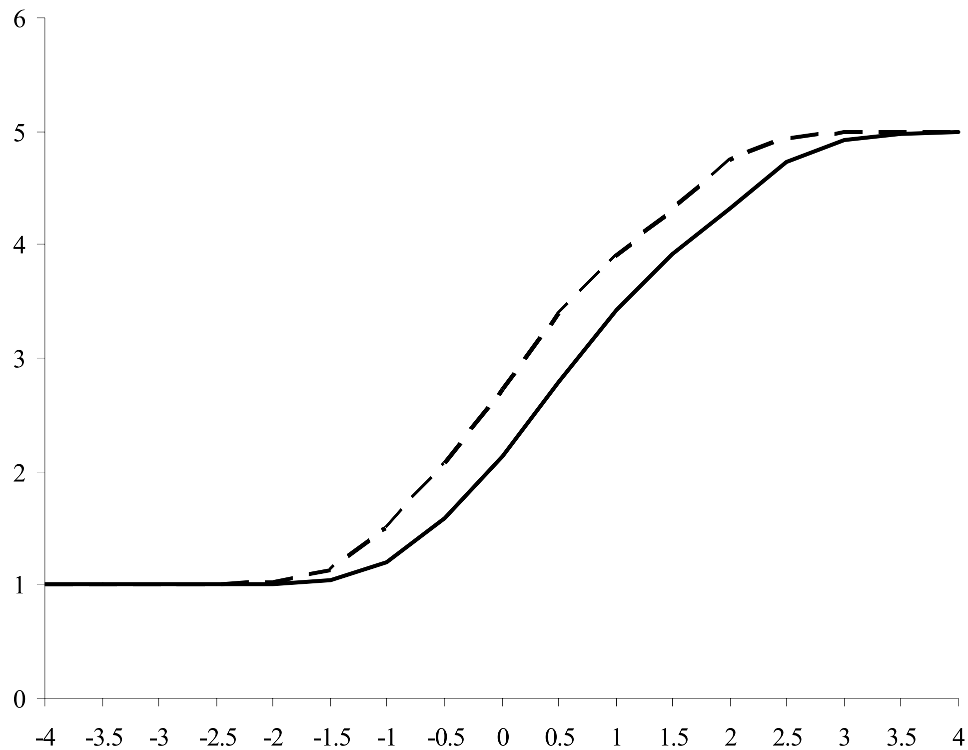


Figure 23 Expected Item Score for Item 53, “Never spend more than I can afford” from the Immoderation subscale. Solid lines represent the age 20 to 30 group. Dotted lines represent the under 20 group.