# GENDER DIFFERENCES IN PERSONALITY: AN ITEM-LEVEL ANALYSIS

by

CASEY C. SMITH

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2020

Gender Differences in Personality: An Item-Level Analysis

by

Casey C. Smith

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

| | |
|---|---|
| _____ | _____ |
| Date | [Charles Scherbaum]<br>Chair of Examining Committee |
| _____ | _____ |
| Date | [Richard Bodnar]<br>Executive Officer |

Supervisory Committee:

Harold Goldstein

Yochi Cohen-Charash

Wei Wang

Zhiqing Zhou

THE CITY UNIVERSITY OF NEW YORK

# ABSTRACT

Gender Differences in Personality: An Item-Level Analysis

by

Casey C. Smith

Advisor: Charles Scherbaum

Personality is a social and organizational construct with a substantial history and discourse. One particular area in personality that is of interest is gender differences in personality. Gender differences have been found on scales measuring various aspects of personality, such as narcissism (Grijalva et al., 2014). While there are differences present in personality data, there hasn't been a consistent explanation for why this occurs. This research looked specifically at the construction of personality items to begin to understand the differences in personality by gender. While social roles and social context are mostly referenced to inform the response patterns of men and women, this research looked to gather greater insight into the impact of social roles and stereotypes on personality items and if the adjustment of the item content can result in the reduction of these differences. To examine these questions, two studies were conducted. The first study explored if certain personality items function differently between men and women. Furthermore, it looked to answer if it is possible to predict which items may show DIF (differential item functioning) by assessing the item for stereotype content. The second study focused on understanding personality item content and the impact of stereotypes through an experimental lens by manipulating personality items to reflect or remove stereotypes and if this would influence the endorsement of the item by gender. The results for the first study indicated

that a small number of IPIP personality items showed DIF. Moreover, a majority of the items that did show DIF were coded to possess stereotype content. There was mixed support regarding endorsement by gender, with the clearest finding that agentic and competent items showing DIF were more likely to be endorsed by men compared to communal, not competent, or neutral items. The results for study 2 showed mixed results as well. The masculine and feminine-written items resulted in the most significant interactions between stereotype, item content, and gender, whereas, the items reflecting the other stereotypes (e.g., agentic, communal, warm, competent) showed fewer statistically significant interactions. The pattern of results in both studies present an opportunity to assess how we measure personality and construct personality items. Psychometric and practical implications are discussed.

**Acknowledgements**

I am grateful to all those that have supported me throughout this journey. My mother, my partner, my sister, my grandmother, my father, my friends and cohort, made me feel it was all possible. Charles Scherbaum, my advisor, guided and provided light to my path. I am appreciative of everyone that took their time to help me grow. I would also like to thank my committee (Yochi Cohen-Charash, Harold Goldstein, Logan Watts, Zhiqing Zhou, and Wei Wang) for their support, feedback, and efforts in shaping this research.

Table of Contents

List of Tables

*Table*

List of Figures

*Figure*

## Introduction

Personality is a research and practice area with a substantial history and discourse (Benet-Martinez et al., 2014). It is a focal construct in organizational research, specifically in the areas of selection, leadership, and motivation (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007; Rossier, 2015). Within this context, personality inventories are often used to determine if individuals are well-matched with the expectations of the job (Barrick, Mount, & Judge, 2001; Hogan & Chamorro-Premuzic, 2015). Over the past 30 years, the use of personality inventories has seen a steady increase in usage (Church et al., 2016). This is not surprising given that scores on personality inventories have been found to predict a variety of organizational human capital outcomes including job performance (Dudley, Orvis, Lebiecki, & Cortina, 2006), task performance (Dudley et al., 2006), training performance (Barrick & Mount, 1991), overall managerial effectiveness, and promotions (Hough, Ones, & Viswesvaran, 1998; Oswald & Hough, 2011).

However, one key consideration and potential challenge with using results from personality inventories in high stakes situations, such as employee selection, is score differences found between men and women, particularly with a lack of understanding of what is causing such differences. Results from different studies and meta-analyses have indicated personality differences amongst genders (Feingold, 1994; Hegelson, 2015; Wetzel, Bohnke, Carstensen, Ziegler, & Ostendorf, 2013). Feingold's (1994) seminal meta-analysis found that men scored higher on assertiveness than women. Women scored higher on anxiety and trust. More recent studies have continued to show gender differences within responses on personality inventories (Grijalva et al., 2015; Wetzel et al., 2013). With these results, a considerable amount of research has focused on understanding these score differences. The explanations that are most often put

forth in the personality research are rooted in biological differences, societal level socialization practices, or national/cultural mores. These explanations represent the underlying assumption that the observed differences amongst genders are real or valid. The differences between men and women are either based on genetics (biological) or develop over time as the individual interacts with his or her environment (socialization, culture). However, an explanation for these personality differences by gender that has been examined less often is the methodology used to measure personality (i.e., the content of the personality items) and how the items in themselves may create or exacerbate score differences between men and women. This explanation posits that measures of personality do not function similarly for men and women. Here, the observed differences are not based on a valid trait differences of personality, but an artifact of the measurement. That is, the measures do not display invariance in the scores across gender groups. Recognizing that this is one possible explanation for the observed gender differences, it is important to add to this area of the research. Exploring the measurement explanation for these differences will provide insights into personality items and potential implications for personality assessments.

Measurement invariance is a statistical indicator that demonstrates that the focal construct of a test or items of a test, in this context, personality traits, are not measured in the same way for individuals who are matched at the same level of the construct but come from different groups. This statistical indicator is often used to examine if a test functions the same for individuals from different groups (e.g., gender, race, or culture). One particular approach to examine measurement invariance is to statistically assess for differential item functioning (DIF; Hambleton, Swaminathan, & Rogers, 1991). The presence of DIF occurs when groups have a different probability of endorsing an item or selecting a particular response even though the groups are the

same on the latent trait/construct, often suggesting that group membership is related to the endorsement of the item or response selection. Within personality, DIF would be present when the probability of endorsing an item reflecting one of the Big Five would be different between men and women, while actually having the same standing on the actual personality factor.

Some studies have found DIF in assessments of personality (Mitchelson, Wicher, LeBreton, & Craig, 2009; Wetzel et al., 2013). These studies found that personality assessment items appear to function differently based on gender, and even race. However, little is known about the factors contributing to gender-based DIF on personality measures, which led to the focus of this research. This research study looked to investigate measurement design and item content factors that are hypothesized to contribute to the presence of gender-based DIF. A key component to this research was to apply theory to predict which items are more likely to show DIF. The two theories central to this prediction are the Stereotype Content Model (Fiske, Cuddy, Glick, & Xu, 2002) and Social Role Theory (Eagly, 1987). Both of these theories were used to derive hypotheses predicting the impact of item content, specifically stereotypic content, on the occurrence of gender-based DIF.

The Stereotype Content Model is a notable theory that organizes and explains the content of stereotypes in society (Fiske, 2012). The model states that stereotypes vary on two core dimensions: warmth and competence (Cuddy et al., 2009; Fiske, Cuddy, & Glick, 2007). Dimensions of stereotypes are comprised of interpersonal and intergroup interactions (Fiske et al., 2002). This model suggests that stereotypes are often mixed with combinations of high versus low warmth and competence. This in turn can create ambivalent stereotypes such that a group can be high on one dimension and low on another dimension. Women are generally viewed as low in competence but high in warmth, whereas men are seen as high in competence

and low in warmth (Fiske et al., 2002; Rudman & Phelan, 2008). The model provides one of the clearest frameworks for understanding gender stereotypes.

Social role theory suggests that sex differences and similarities in behavior result from gender role beliefs, which represent perceptions of men's and women's social roles in their specific societal context (Eagly & Wood, 2001). Gender role beliefs come from the observation of gender-associated behavior. Inferences are made that the sexes possess corresponding dispositions, such that men and women are believed to have attributes that are matched for sex-typical roles. These sex-typed attributes then become gender stereotypes or consensually-shared beliefs (Eagly & Wood, 2001). This theory provides insight into how stereotypes are integrated into everyday expectations and socialization.

Both the Stereotype Content Model and Social Role Theory are elemental to investigating if the embeddedness of stereotypes in personality assessment items leads to measurement error such that the item is not perceived similarly by different genders. The general hypothesized direction of this research is that measurement invariance is occurring because some items on personality inventories may possess content that is similar to commonly held societal gender stereotypes. To test this hypothesis, two studies explored applying a theoretical explanation, focused on gender-based stereotypes, to the presence of measurement invariance. The first study looked to understand the content of personality items that are expected to show DIF. This study included coding archival personality data based on the stereotype content model (Fiske et al., 2002) and social role theory (Eagly, 1987). The second study expanded Study 1 by examining if score differences can be experimentally induced by manipulating the stereotypical content of personality items to induce a gender-associated stereotype and it assessed the impact on the response patterns by gender.

The value of this work is to dive deeper into the personality assessments that are now often included in employee selection process. Moreover, this research looks to examine how the construction of an item can be influenced by social elements, such as stereotypes and social roles.

## Personality as a Construct

Personality is defined as the unique combinations of the attributes, qualities, and characteristics that distinguish the behavior, thoughts, and feelings of individuals (Pervin, 2001; Saucier & Srivastava, 2015). Personality is considered to be a set of psychological traits and mechanisms within a person that are organized, enduring, and influence interactions with the environment (Larsen & Buss, 2005). Additionally, while determining what personality truly is, research has also focused on how personality is structured. Personality is a complex construct and one key component of understanding it lies in delving into its theorized structure.

The trait theory describes how personality is conceptualized (Mischel, 1996). Trait theory was developed in the 1940s and 1950s by Cattell and Eysenck (Pervin, 1994). The trait theory/approach has arguably been the most dominant theory used to describe personality. This approach states that human behavior is either rooted from an underlying trait or internal processes that predispose the individual to engage in certain behaviors.

The basis of the trait theory is that it infers the underlying personality structure and compares persons and groups on trait dimensions (Mischel, 1996). The trait approach also posits that behavior is relatively stable, as traits are highly consistent across situations (Fleeson & Jayawickreme, 2014). The value of this theory is that it directly connects the presence of a trait and its degree to the personality of an individual. Regarding the trait theory, the definition and construction of the trait profile is also a key component.

As early psychologists attempted to define personality and develop personality theory, traits acted as the fundamental unit of personality (Winter & Barenbaum, 1999). Allport (1921; 1996) defined traits as motivational personal dispositions. The basic assumption is that traits are attributes that are common among many people and as a result, behavior can be classified and generalized. Traits serve as indicators to summarize, predict, and to explain a person's conduct (John & Gosling, 2000). As individuals perceive their surroundings, they are cognitively inclined to encode and categorize patterns. Patterns allow for the prediction of future behavior, so individuals seek out this information (Kelly, 1955). In relation to perception of people, these patterns are classified as traits. Moreover, traits connect the explanation for behavior to the individual rather than the situation, reflecting that there is an internal process at work (John & Gosling, 2000).

The second component of the trait profile is its construction or how many traits make up personality. The number of personality traits used to describe an individual's disposition has been an area of active research for some time. After the foundation of traits from Allport, Cattell and colleagues went a step further and used the inductive-hypothetico-deductive factor-analytic method, which is an exploratory factor analysis to identify items that reflect personality traits (Boyle & Helmes, 2009). As personality developed, numerous trait structures were considered but the field has focused on five (Digman, 1990; Goldberg, 1981). For the last thirty years, researchers have found consistent evidence that personality traits can be organized into a hierarchical structure with five traits being the core of describing a person's personality (Fleeson & Jayawickreme, 2014). This structure is known as the Five Factor Model, which will be discussed in further detail and due to its prominence will be the focal personality model used in this research.

The Five Factor Model is regarded as the universal description of personality (Costa & McCrae, 1997, 1998, 2008, 2009; Digman, 1990; Goldberg, 1990). This model was inductively and empirically created (McCrae & Costa, 1996). It was derived from factor-analytic studies on personality trait data and became the field's consensus (Hough & Connelly, 2013). The basis of the five-factor structure is reliant on the natural language (Goldberg, 1993). More specifically, the model was created from analyses of everyday language that people use to describe themselves and others and was supported by empirical analysis (John & Srivastava, 1999). The five factor traits were pulled from the dictionary, whereas all words that were considered personality related was of focus (John, Angleitner, & Ostendorf, 1988; Saucier & Goldberg, 1996). This inherently means that the words used to label personality traits are embedded in natural language and are considered significant and useful for communication (Goldberg, 1981).

The five classically described traits in the five-factor model are openness to experience, conscientiousness, agreeableness, extraversion, and neuroticism (McCrae & Costa, 1997; 1999). Openness to experience is regarded as a need for variety of experience, novelty, or change. Characteristics closely linked to this trait are imaginative, cultured, intelligent, and original (Barrick & Mount, 1991). Conscientiousness is defined as achievement striving, a strong sense of purpose, self-discipline, and high aspiration. This trait reflects dependability, carefulness, being thorough, responsible, and organized (Barrick & Mount, 1991). Agreeableness is considered as compliance or a willingness to defer to others, particularly during interpersonal conflict. Extraversion is often characterized as gregarious, assertive, active, and sociable (Barrick & Mount, 1991). Lastly, neuroticism is a tendency to experience dysphoric affect, like sadness, hopelessness, and guilt (McCrae & Costa, 1997; John & Srivastava, 1999). It is considered a broad domain of negative affect (Costa, Terracciano, & McCrae, 2001). Associated

characteristics of neuroticism are anger, being anxious, emotional, insecure, and worried (Barrick & Mount, 1991). Each of the five broad factors has a number of sub-factors. The sub-factors are specific facets or aspects that branch off the domains and categorize personality trait dispositions (Christiansen & Tett, 2013). These sub-factors also account for additional variance of personality above the higher-level factor. Table 1 presents the list of factors and subfactors of the five-factor model.

The Five Factor Model's primary addition to the field is that it reflects personality as a commonality among all persons and provides a hierarchical structure. The advantage of its structure is that there is a universal understanding of the words that define the factors and the language does not ascribe to a certain theory (John & Srivastava, 1999). The Five Factor Model has been found to generalize across languages, cultures, raters, and samples (John & Srivastava, 1999). With this universality, the construct of personality can be applied to everyone. Personality coming from the lexicon of common language allows for any individual within the society to be categorized and classified by it. Moreover, these five factors allow behaviors to be categorized. The hierarchical structure provides order to a large selection of trait concepts and provides a framework for researchers to identify similarities and differences among other models of personality structure (John & Gosling, 2000). In addition to the structure of the Five Factor model, the way in which the model is developed is integral to its application.

**Development of the Five Factor Model**

The development of the Five Factor Model is a significant component of its application in other areas of work. As previously mentioned, this model was created from factor analyses of everyday vernacular and was described as natural language constructs (Allport & Odbert, 1936; Goldberg, 1990, 1993; Tupes & Christal, 1992). To develop this, Allport and Odbert (1936)

generated a list of words from the English language that were considered to describe personality. Allport continued this by using synonyms to further develop the ecosystem of personality traits. Following Allport, the factor analytic approach was used by psychologists to continue to construct and develop the model of traits (Thurstone, 1934; Cattell, 1947). The common denominator of all these early approaches is the lexical nature of the words that describe personality. What is key here is that the way we label people's dispositions is predicated on how we talk in everyday conversation.

Personality as a construct stems from the social environment and the language that exists within that environment. An important criterion of a valid personality theory is that the factors of the model are socially relevant (Eysenck, 1991; Oswald & Hough, 2011; Roivainen, 2013). As personality psychology research grew, the reliance on "everyday" language to assess personality became the standard methodology and source of data (Baumeister, Vohs, & Funder, 2007; Matthews, Deary, & Whiteman, 2003; Schwarz, 1999; Uher, 2013). As such, the commonality and universality of personality traits is sourced from language. This concept can be described as the lexical hypothesis. The lexical hypothesis is a framework that explains the interaction of words in the social environment and its influence on personality. Language is inherently social and shaped by the environment it exists within. Language is also embedded in how society operates; it is a key component of defining social roles.

The foundation of the lexical hypothesis was first recognized by Galton in the late 1800s. Galton proclaimed that individual differences between humans would be encoded into single terms across all the world's languages (Goldberg, 1993; Uher, 2013). Over time, socially shared constructs of self and other representations are encoded into the human language (Uher, 2013). People then encode salient and socially relevant differences between individuals from

conversational, everyday language. The lexical hypothesis suggests that personality traits are embedded in the English language (Allport, 1937; Cattell, 1946; Saucier & Srivastava, 2015) and a valid personality inventory identifies personality concepts within a linguistic and cultural context (Saucier & Srivastava, 2015). Language is a key component in the way personality traits are classified and grouped into factors. Two fundamental principles of the lexical hypothesis are that personality characteristics that are significant to a group will become a part of the group's lexicon. The second principle states that the most significant characteristics are most likely to be encoded in its simplest form (John, Robins, & Pervin, 2008).

Overall, the lexical hypothesis is a prominent framework to the construct of personality. It provides a deeper dive into how personality is weaved into the way we speak or classify ourselves and thus the construction of personality assessments. The reflection of all the social influences and implications are also embedded in how personality is explained, interpreted, and presented. While trying to dissect the results of personality assessments, the consideration of the source of personality items is key. As the construction of personality traits are based on the lexicon used in society, other social constructs may also be present in the items and interpretations from personality assessments. Delving into assessments of personality, the interaction of the social aspects, like social roles and stereotypes, and the measurement of personality itself is critical to our clear understanding of how we conceptualize personality.

The Five Factor Model has been found to be a universal model of personality. However, findings have shown variations outside of individual-based differences. One distinct area where discussion exists about the presented differences in personality is when looking at gender. Some research suggests that the differences found between gender reflect true differences between men and women. Other studies have questioned if the patterns found based on gender are due to

construct irrelevant factors. Examining the interaction of gender and the Five Factor model provides context into the larger purpose of this research, which is to explore the impact of personality assessments, particularly when looking at differing scores based on gender.

**Gender and the Five Factor Model**

Research has documented the mapping or connection of gender and personality traits, particularly with a linkage to the prominent five factors. The impetus for looking at gender differences in personality research began with Maccoby and Jacklin (1974)'s review of sex differences in cognition, temperament, and social behavior (Feingold, 1994). They found that on personality inventories, men tended to show scores that indicated that they were more assertive or dominant, more aggressive, and less anxious. This work provided an establishment of interest in these differences.

An entire subsection of the personality literature investigates the inferences of gender and personality on outcomes in various contexts. Within the literature, gender differences effect sizes range from a quarter standard deviation to a half of a standard deviation (Costa et al, 2001; Lockenhoff et al, 2014; Hyde, 2005). There are numerous attributions to explain the variability of gender differences found: the subfactor level versus the trait level, different types of personality assessments, methodological limitations impacting the personality score, or that the differences between the sexes are close to zero (Jaffee & Hyde, 2000; Soto, John, Gosling, & Potter, 2011).

Assessing findings by trait level has been one of the stronger focal points in personality research. Some results indicate gender-based personality differences happen at the broader trait level. Looking at neuroticism, women were found to score higher than men on neuroticism consistently, an effect which was replicated across 37 countries (Lynn & Martin, 1997). Scores

on the neuroticism scale for women are .25 to .50 standard deviations higher than the average of men scores (Escorial & Navas, 2007). On the traits of agreeableness, extraversion, and conscientiousness, the differences are more generally found to be a less robust difference between men and women (*d* =.20; Soto et al., 2011). In Feingold's (1994) meta-analysis, women were found to be slightly higher in extraversion than men. However, a different meta-analysis found that men generally have higher scores on extraversion than women (Lynn & Martin, 1997). Lastly, conscientiousness hasn't yielded large or consistent differences between men and women (Costa et al, 2001; Feingold, 1994; Stake & Eisele, 2010). In summary, gender differences at the higher order facet level are consistently present, but the effects are often less than a half of a standard deviation and closer to one-quarter standard deviation (Costa et al, 2001). This pattern of findings at the facet level has led researchers to the further investigation of differences by gender in personality by sub-factor.

Research at the subfactor level provides some insight into where gender differences in personality are presenting themselves. Looking at sub-factors of neuroticism, Feingold (1994) found that women have higher scores on anxiety (*ds*= .26 to .32) and depression scales. On the sub-factor of hostility, there is varying findings with women sometimes being higher, men being higher, or no difference at all (Averill, 1983; Ross & van Willigen, 1997; Scherwitz, Perkins, Chesney, & Hughes, 1991).

Delving into sub-factors of extraversion, it is often posited that men and women will show differences on subfactors based on the expressive nature of the trait (Stake & Eisele, 2010). Women were found to score higher in warmth, gregariousness, and positive emotions, which are the more expressive traits, but lower in excitement-seeking and assertiveness than men (Feingold, 1994; Costa et al., 2001). Men tend to score higher on measures of agreeableness

compared to women, even at the sub-factor level (Costa et al., 2001). In the Feingold (1994) meta-analysis, men were lower in trust and nurturance (*d*s= -.25 to -.28) On the trait of openness to experience, men scored lower in openness to aesthetics, feelings, and actions, but higher in openness to ideas. There were no consistent differences by gender on openness to fantasy or values (Costa et al., 2001). Lastly, dutiful and order, two subfactors of conscientiousness, were shown to be more prevalent in women (Costa et al., 2001; Feingold, 1994). The reported mean differences (z-scores metric) for each sub-factor from Costa et al. (2001) is presented in Table 2.

Overall, neuroticism and agreeableness show clear differences by gender on the subfactors; the other three factors (extraversion, openness, conscientiousness) have conflicting differences on their subfactors. This makes the consistency of gender differences on personality traits a bit unclear. However, this pattern does show that at the sub-factor level, some gender differences are more apparent.

A compilation of all these results summarizes that there are discrepancies in the findings of personality based on gender. As the gender differences in personality are varied by sub-factor, there may be other patterns of differences when looking one step deeper at the item level. Before investigating a measurement explanation, it is important to examine the traditional theories that the literature has mostly attributed gender differences to, which are biological, evolutionary, and social. All of these explanations are considered to be more stable determinants of behavior. However, understanding the fluidity of gender-based personality results, there is some disconnect between the results and the stability of these explanations.

**Explanations of Gender Differences in Personality**

Feingold (1994), the seminal meta-analysis in this research, referenced three prominent explanations for gender differences in personality - biological, evolutionary, and social. These

approaches are historically referenced when reflecting on the differences based on gender in personality findings. However, the variation of gender differences found in the research puts into question these more stable explanations. In addition to exploring these explanations, this research will introduce a fourth explanation which is focused on the *measurement of personality*. How personality is measured has not be as frequently discussed but may provide critical insight into the pattern of gender differences that have been found.

***Biological***

Early researchers believed that trait differences were biologically based (Fausto-Sterling, 1985; Feingold, 1992). Genetics, hormones, and brain structure and function are the basis of the biological explanation of gender differences in personality (Hegelson, 2015). When looking at gender differences, the thinking behind the initial research was that evidence of such differences supported that individual differences in traits were biologically rooted (Fausto-Sterling, 1985; Feingold, 1992). Twin studies were used to measure if personality traits were heritable. Through the lens of personality being genetic, this suggests that characteristics that individuals' exhibit are innate to the person and can be considered dispositions (South et al., 2015). Dispositions are described as enduring patterns of emotionality, self-regulation, and orientations to the social and physical environment that characterizes the individual (Donnellan et al., 2015). The biological explanation certainly provides insight into some personality differences, however, it does not fully account for the gender differences we see in personality. This explanation would suggest more stability in personality differences between genders. Since gender-based differences that have been reported are not found consistently, the biological explanation may not fully explain these differences.

***Evolutionary***

A second explanation of personality differences by gender is based on evolution. This theory suggests that females and males possess dissimilar traits due to different evolutionary needs, primarily to maximize reproductive success (Buss, 2007). This explanation conceptualizes that there is a variation of adaptations (or evolved psychological circuits) that produces individual differences (Buss, 1995; Buss & Penke, 2015). Here, personality traits are considered functional strategies that individuals use to solve specific problems during evolution (Buss, 2009). The adaptations for men and women are expected to differ in situations where they may recurrently encounter adaptive problems over history (Buss, 1995). This explanation also states that in contexts where adaptive problems are the same, it is surmised that both men and women will be psychologically similar.

Buss and Penske (2015) stated that the largest effects of gender differences are present on the trait of neuroticism. The evolutionary hypothesis for this is that social threats have been costlier because of a women's responsibility for childcare. Within this same explanation, high neuroticism in men would have prevented them from taking risks which was considered mandatory for competition (Buss & Penske, 2015). The social situation created a circumstance where a difference between genders manifested. Research in this area focuses on identifying the domains that consisted of different adaptive problem for men and women. The evolutionary explanation is one that is insightful and provides a background in how gender differences may have been manifested in society. However, with what we know about the variation in personality results, the distal explanation of evolution may not fully account for all the differences and the variation of those gender differences in personality across the literature.

***Social***

Transitioning to a more proximal explanation, social theory is one that is examined to account for the differences that have been reported between genders. Social environmental theories look to the social environment to explain how gender-related traits are developed (Hegelson, 2015). The prominent theories are social role theory and gender-role socialization (Hegelson, 2015). These theories share a common thread, the impact of social construction on gender.

The social role theory explains that the roles in society dictate or guide how men and women behave (Costa et al., 2001; Eagly, 1987; Eagly & Wood, 1991). Social roles are markers and/or shared expectations of what behavior is typical, expected, and deemed appropriate in the overall context or environment. Specific characteristics are aligned to certain social roles. These characteristics are described as agentic and communal (Table 3). Agentic characteristics are classified as self-assertion, independence, aggressiveness, and, mastery. Communal characteristics are represented as selfless, concerned with others, sensitivity, sympathy, and maintaining relationships (Eagly & Steffen, 1984; Eagly & Johannesen-Schmidt, 2001). It is perceived that men are directed towards achieving agentic based goals and women are directed towards achieving communal-based goals (Bem, 1974; Block, 1973; Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Spence & Helmreich, 1978). These goals suggest that men are motivated to act in line with agentic behaviors and the opposite for women.

The manifestation of agentic and communal-associated roles is based on two prominent observed differences in society (Eagly & Steffen, 1984). The premise here is that if the behavior is observed, it is more likely to believed that the actions are specific to that group. One societal observation is that women are more likely placed in roles that are at low levels of status and

authority, whereas men are often at the higher end of the hierarchy and level. The observed role distribution is that women are more prominently in the home and men are often in the employed workforce. As a result of these common placements of men and women, it lends to the continued stereotypic belief or ascription that women possess personality characteristics that make them more communal than men and that men possess personality characteristics that make them more agentic. This view expresses that the prominence of actions by groups witnessed in social settings is an influence to the personality attributed to these groups.

Another social theory that explains the impact and influence of the social environment on expressed gender differences is gender socialization theory. Gender socialization theory examines how men and women learn masculinity and femininity and how they are socialized into traditional gender roles (Carter, 2014). The inputs of socialization can come from culture (race, ethnicity, language, play, competitive sports, visual and print media), institutions (school, religion, workplace), parenting style, and peers (Philpot, Brooks, Lusterman, & Nutt, 1997). The socialization process is hypothesized to start even at birth with parents treating infants differently based on the sex (Messner, 1992). Children also identify with their same sex parent and this identification is the process where children incorporate their parents' gender role behaviors into their own self (Freud, 1964). Socialization explained in the social-cognitive perspective states that gender role behaviors are learned through reinforcements such as rewards and punishments and observational learning (Bandura, 1977; Bem, 1974; Eccles, Jacobs, & Harold, 1990; Lott, 1994). With reinforcements, society, including parents, rewards children for acting in a manner appropriate to gender roles and these rewards inform expectations. For example, women learn not to be assertive in interactions with men. Men learn that they can engage in risky behavior (Eccles, 2001). Social-cognitive theorists suggest that as children reach adolescence it is only

then in which they have the cognitive capacity to create a schema that is different from societal expectations (Eccles, 2001). Additionally, they must have had opportunities to observe examples of gender role transcendence and encouragement by their social surroundings (Eckes, 2000). If one is to live in a society that has clear gender roles and strict rules, then it is a greater likelihood that the established gender norms will not be questioned. However, if the society is more egalitarian in its gender role prescriptions, there is room for less rigidity in gender roles and an openness to variation in gender subscription (Eccles, 2001).

When considering the social explanation of personality differences, it is important to remember that this viewpoint considers these differences true differences (i.e., the differences between genders are real). This explanation suggests that the environment shapes men and women tendencies, influences their behavior, and is the source is the specific variation of traits (Hegelson, 2015). Similar to the biological and evolutionary explanations, it is presumed that exhibited gender differences are reflective of true, valid characteristics of men and women. However, the variations in results reflecting gender differences suggests the social explanation does not fully account for the gender differences we see in personality. These discrepancies make it hard to argue that there are consistent and stable differences. If these differences were explained by biological, evolutionary, or social theories, one would expect more consistency in the differences across studies. Shifting away from explanations tailored to inherent differences between men and women, the measurement explanation considers item construction of the personality assessment as a predictor of the differences found between men and women is discussed.

***Measurement***

Measurement is not as often included in the traditional theoretical explanations of how gender differences manifest or appear in the personality domain. The measurement viewpoint targets the way personality is measured as a critical component to the interpretation of the results of the assessment. Measurement as a factor is different from the biological and social explanations because it allows for the exploration of the impact of variables that are not internal to the individual (i.e., the differences are not necessarily real). Furthermore, the way in which constructs are measured can influence the data that follows, as such, the results may be representative of multiple aspects and include construct irrelevant variance. For the results of an instrument to be found legitimate, it is important that the instrument functions equivalently across the groups that are being compared (Wetzel et al., 2013).

The specific focus of establishing the measurement explanation is to ascertain if results from personality assessments regarding gender differences are capturing construct irrelevant variance. Construct irrelevant variance is considered present when an assessment contains variance associated with outside constructs, contamination, or other factors unrelated to the construct (Cronbach & Meehl, 1955; Messick, 1995). These sources of variance may be related to other distinct latent traits or method related variance (i.e., response pattern). To determine if construct irrelevant variance is present, statistical analyses are used to delineate what is happening methodically. One statistical characteristic that is used to find if there is a presence of construct irrelevant variance is differential item functioning (DIF).

Differential item functioning occurs when items are found to function differently across groups that are matched on the construct. Looking at the item level, an item is considered to possess DIF if test takers from different groups have equal ability, but have an unequal probability of item success or endorsement. The presence of DIF represents that the measure

may have some additional, irrelevant variance which is impacting the scores differentially between groups. Within this paper, differential item functioning is explored as a source of gender differences on personality items.

When looking at other research domains, measurement has been included as an explanatory factor for results. There has been an investment and a critical view on how measurement instruments are constructed, employed, and interpreted. Extensive efforts have been taken in understanding the calibration of measurement, particularly in attitudes and cognitive ability (Sackett, Lievens, Van Iddekinge, & Kuncel, 2017). Using cognitive ability as an example, research efforts have investigated the different ways in which intelligence is measured and how those modes of assessment may impact the results (Goldstein, Scherbaum, & Yusko, 2009; Hough, Oswald, & Ployhart, 2001). The history of cognitive ability or intelligence shows attempts to validly assess the construct (Fleishman & Quaintance, 1984; Ghiselli, 1966; Hunter & Hunter, 1984; Landy, 1989; Wasserman & Tulsky, 2005). Research in the domain of cognitive ability is illustrative of the way measurement affects the results or the interpretations of the findings. This work provides a framework to explore how external variables impact the results of measures, like personality assessments. A deeper dive into measurement has not been as developed in the domain in personality, and even more specifically when applying it to the interpretation or explanation of gender differences as they appear in personality, but some work has been done.

Few studies have examined the linkage of measurement to gender differences found on personality assessments (Smith & Reise, 1998; Mitchelson, Wicher, LeBreton, & Craig, 2009; Reise, Smith, & Furr, 2001; Smith, 2002). These studies in general questioned if gender differences are valid or if they are attributable to outside factors (e.g., the properties of the

assessment; Smith & Reise, 1998; Reise, Smith & Furr, 2001). Smith and Reise's (1998) found that the mean differences by gender were somewhat affected by the items functioning differently. Reise, Smith, and Furr (2001) found that 33% of the personality items on the Big Five personality trait taxonomy showed DIF. Many of these items were on the Anxiety subscale and favored either men or women. Wetzel et al. (2013) analyzed the German NEO-PI-R to see if the instrument was functioning unequivocally between men and women. The authors investigated if there was a potential confounding variable in response styles, looking particularly at if extreme versus moderate response styles impacted endorsement of personality traits by gender. Their research found that some items in each of the five facets showed differential functioning by gender, such that certain items had a higher probability of men endorsing them and other items had a higher probability of women endorsing them. They found 17 items favoring men and 14 items favoring women (Wetzel et al., 2013). Examples of the items that were operating differently for men and women were "I am easily frightened (Neuroticism)" and "I am easygoing and unconcerned (Conscientiousness)." They additionally looked at the impact of an individual's proclivity to respond to questionnaire items, or response styles. Here, it was found that controlling for response style had a small effect on the items that showed differential functioning. The way people responded did not impact the different ways the items were endorsed by men and women. These studies within the last decade have worked to identify if differential item functioning can explain the observed gender differences on personality measures.

Finding DIF on personality inventories is critical because the presence of DIF may suggest that test scores or findings are no longer comparable between groups (Wang, 2008). As research has found some evidence of DIF, much of it still has not identified or pinpointed the

source. The question that has remained largely unanswered is why is DIF occurring in personality assessments. Outside of additional statistical techniques, it is critical to look at the way personality items are constructed and what may impact the presence of DIF. Instead of simply identifying that items show differences in endorsement, these studies will add value by also predicting why and which items are showing differences by gender. This research offers hypotheses that stereotypes embedded in personality items are a potential explanation of the invariance that has been found on personality inventories. Furthermore, this research establishes a theoretical basis that can be used to predict which items are resulting in measurement error.

## Sources of Gender-Based Measurement Invariance: Stereotypes

The foundation of this supposition that stereotypes are embedded in personality items is that stereotypes and personality share a common feature, language. In order to offer hypotheses regarding the impact of stereotypical content in personality items, it is important to review how stereotypes are developed and organized through language. To determine if there is an impact of stereotypes in the endorsement of the items, the construction of stereotypes and two models (Stereotype Content Model and Social Role Theory) are discussed. Starting with the construction of stereotypes, we will examine how stereotypes are intertwined in language and transmitted through the lexicon.

### *Stereotypes and Language*

The connection of language to stereotypes and personality is interesting as Allport was an early researcher and theorizer in both areas. Language is a simple mechanism where people are categorized and how stereotypes are shared (Allport, 1954; Fishman, 1956). Allport (1954) explained that linguistic terms define content and also act as organizing principles and evaluative references. His early observations noted that different groups have contrasting stereotypes.

Stereotypes are defined as consensually shared, developed beliefs about the characteristics of social groups (Stangor & Schaller, 2000). These beliefs are stored as information, encoded in memory, and then are retrieved to direct associated responses when interacting with others. In the context of language, stereotypes are defined as mental associates between category label distinctions and traits. Stereotypes are learned and maintained through language and culture. Overall, language is how stereotypes are defined, communicated, and assessed (Maass & Arcuri, 1996).

Language is a critical component to stereotype transmission, cognitive organization, stereotype maintenance, and expression of stereotypic identities (Maass & Arcuri, 1996). Language fosters stereotype transmission by being the mode of how stereotypes are transferred between persons and across generations. Stereotypes are embedded in the vocabulary of a given language. Then, those relevant social beliefs are embedded in the lexicon and organically absorbed as language is learned and acquired.

In reference to cognitive organization and stereotype maintenance, language provides key terms (e.g., stereotypes) that information is organized around. One example of these key terms are social category labels which act as anchors for other reference points and influence how stereotypic information is grouped. This grouping function is also present in personality. Personality traits and related behavior are connected to a label and a network of associated terms (Stangor & Lange, 1994). When the label is activated, conscious or unconsciously, the full network of associated terms, behaviors, and characteristics is also activated. The activation process for category-based concepts and traits is more often automatic, such that, stereotypes can be quickly and subconsciously referenced.

When looking at stereotype maintenance, language is also used as a mechanism to work against disconfirmation of existing stereotypes. This process interacts with the dichotomy of stereotypes usually expressing favorability for the ingroup and unfavourability for the outgroup. Looking at traits and the process of disconfirmation, in an experiment conducted by Rothbart and Park (1986), it was found that unfavorable/negative traits needed fewer instances of confirmation, and once confirmed, it took more efforts to disconfirm and negate those unfavorable traits. Outgroup stereotypes, which are mostly comprised of negative traits, are resistant to change because negative traits require more to be disconfirmed (Maass, Montalcini, & Biciotti, 1998). Related to the disconfirmation aspect, language also maintains stereotypes through the scope of personality traits. Personality traits can be broad or narrow. Research has shown that broad traits are used for positive characteristics of the ingroup and used for negative characteristics of the outgroup (Hamilton, Gibbons, Stroessner, & Sherman, 1992). In summary, language supports stereotype maintenance in simply how they are constructed and the impact of how we speak to the behaviors of others. Negative language is much harder to disconfirm and different scopes of language are associated with particular characteristics of groups. Much of this literature is similar to the social identity theory (Tajfel & Turner, 1986) such that there are motives to protect and/or enhance one's social identity. In this case, it is done through linguistic means.

Overall, language is integrated into the construction and maintenance of stereotypes. More importantly, when looking at both stereotypes and personality, language is a commonality. The properties that are shared make these constructs (stereotype and personality) more linked than has been previously examined. Understanding this linkage and connection, this research will examine stereotypes as a source of DIF in personality scores. Personality is linked to language,

and thus stereotypes, which would suggest that stereotypes are present when personality is measured. As such, stereotypes are likely to be present in personality items. In the following section, two prominent models of stereotypes are described and hypotheses about how gender stereotypes lead to DIF are offered.

## Stereotype Models and Gender-DIF on Personality Items

Stereotypes have a large body of research that explain why they exist and how they operate. The literature is expansive, including, racial, ethnic, religious, cultural, age, and gender stereotypes. The two most prominent models for gender stereotypes are the Stereotype Content Model and Social Role Theory. Both theories provide insight into how gender stereotypes operate and their implications for perceptions of men and women.

### *Stereotype Content Model*

The Stereotype Content Model is based on the observation that stereotypes are often stable and have a systematic structure and foundation. In this model, stereotypes are conceptualized as having two dimensions, warmth and competence. Warmth and competence are viewed as the universal and fundamental pieces of social perception (Cuddy, Fiske, & Glick, 2008; Fiske et al., 2007). The combination of these dimensions results in different intergroup emotions (e.g., pity, envy, contempt). Warmth is the dimension in which people use to decide if another person is friendly, trustworthy, or honest. Competence represents the focal group's perceived ability to be successful at tasks considered important (Eckes, 2002). Warmth signifies the other person's intent and competence signifies the other person's ability to pursue the intent (Fiske, Cuddy, Glick, & Xu, 2002). Moving to the guiding force behind stereotypes, the Stereotype Content Model states that two social structural variables, status and competition, predict the dimensions of stereotypes. Societal status is hypothesized to predict perceived

competence and competition predicts perceived warmth. Research has shown that when there is an interpretation of others, warmth is judged before competence and has a greater effect on how a person is viewed (Fiske, Cuddy, & Glick, 2002). Warmth has been a component even from early studies from Asch (1946). Applying this to men and women, it is often a mixed stereotype. In general, women are perceived as low in competence, but high in warmth, and men are viewed as high in competence, but low in warmth (Eagly, Wood, & Diekman, 2000; Eckes, 1994; 2002).

The Stereotype Content Model is a two-dimensional, diagonal that places different groups in four stereotype categories: low warmth/low competence, high warmth/high competence, low warmth/high competence, high warmth/low competence. Warmth and competence have been each broken into two specific sub-components. Warmth includes emotionality and empathy and competence includes dominance and efficiency (Linssen & Hagendoorn, 1994). The warmth component is associated with traits such as morality, trustworthiness, sincerity, kindness, and friendliness (Cuddy, Fiske, & Glick, 2007). Competence has been synonymous with capability, skill, and talent (Kervyn, Fiske, & Yzerbyt, 2015). Traits associated with competence are clever, competent, creative, efficient, foresighted, ingenious, intelligent, knowledgeable. From Fiske et al.'s (2002) research, the traits used to measure competence on a scale were "capable, efficient, skillful, competent, confident, and intelligent." The traits used to measure warmth on a scale were "well-intentioned, warm, good-natured, friendly, trustworthy, and sincere." Following this, many research studies have continued the application of this model and the traits that are associated with them (Gaucher, Friesen, Neufeld, & Esses, 2017; Rast, Gaffney, & Yang, 2017; Johnson, Stevenson, & Letwin, 2018).

An additional component to the structure of stereotypes is the expectation of behavior or perceptions of individuals that comes with stereotypes. Along with the development of the

Stereotype Content Model, Glick and Fiske (1996) also developed the theory of ambivalent sexism, which explains the ways that women can be perceived based on their endorsement of the stereotypes associated to women. Ambivalent sexism is the combination of hostile and benevolent sexism. This is important as it provides some insight into how endorsement of stereotypes by all groups can be influenced.

Both versions of sexism encourage a traditional view of women. Hostile sexism posits that negative attitudes and stereotypes are held towards nontraditional women. It is based on the belief that men should have more power than women, that women are inferior to men on competence-related traits, and that women's sexuality is threatening to men's status and power (Eckes, 2000; Glick & Fiske, 1996). Benevolent sexism predicts positive attitudes and stereotypes toward traditional woman (Fiske, Xu, Cuddy, & Glick, 1999; Glick, Diebold, Bailey-Werner, & Zhu, 1997). This version of sexism is the belief that men should protect and provide for the women they are closely connected to and dependent on, that women are the superior gender in the context of conventional gender roles, and that men can achieve happiness when romantically involved with women (Eckes & Trautner, 2000; Glick & Fiske, 1996).

When looking at women's acceptance of these views, women may often endorse benevolent sexism, especially in a culture that is more sexist (Glick et al., 2000). Additionally, benevolent sexism does not have the punitive consequences as hostile sexism. Hostile sexism acts to penalize women who do not conform to acceptable, gender-guided roles. Benevolent sexism, in contrast, states that women are rewarded when they conform. Women who embrace the conventional roles are valued and protected (Glick & Fiske, 1997). The combination that describes traditional women is low status, warm but incompetent. Nontraditional women are described as high status, no warmth, but competent. When discussing all of these descriptions,

there is the recognition that they are based on expectations of behavior. The presence of sexism and how it operates helps to dive deeper into the purpose of this research, such that there are influences and concepts that could be impacting how stereotypes are viewed and endorsed. Additionally, if stereotypes exist within personality items, this strength of gender-based expectations provides insight into the impact on personality assessments and endorsement of personality items.

The Stereotype Content Model provides a systematic, theoretical framework to map personality traits to the language of stereotypes. This framework also allows for the prediction of which personality traits are also linked with stereotypes and how it is a part of the measurement of personality. Another theory that is instrumental to understanding stereotypes and the predictive framework to this research is the social role theory. The social role theory was briefly referenced in the social explanation of gender differences. The additional elaboration and inclusion of this theory here reflects how social structures are also intertwined in stereotypes.

***Social Role Theory***

Alice Eagly and her colleagues developed a body of work on the nature of gender stereotypes and this work has cumulated into the social role theory. The social role theory explains that the inferences made based on the observation of women and men's role-constrained behavior are reflected in the gender stereotypes about men and women's dispositions (Eagly, 1987; Eagly, Wood, & Diekman, 2000). As each gender is more associated to particular role in society, inferences are made about individuals' disposition based on their gender. The start of this association occurs as men and women are prepared for certain roles by being socialized to possess personality traits and skillsets that support the associated role performance (Eagly & Wood, 2012). This influences social regulation to meet other's expectations about men and

women, and extensive socialization to promote personality traits (Eagly & Wood, 2012). This theory also illustrates that there are consequences to not endorsing or engaging in gender-based stereotypes. Men and women are rewarded for conforming and penalized for deviating from stereotypical behaviors (Eagly & Wood, 2012). In Eagly's (1992) meta-analysis, women were given negative evaluations when they acted more aligned in a male-stereotypic assertive style. This theory provides direction to what can drive men and women to behave in alignment to their gender-based expectations.

Similar to the stereotype content model, the social role theory presents insight into what impacts stereotypes and how they are constructed. This theory explains the force behind the stereotypes that are intertwined in the lexicon. Social role theory is dependent on gender roles. Gender roles are shared by society through language and socialization. In this theory, the warmth and competence dimension are more commonly labeled agency and communion (Bakan, 1966; Eagly, 1987; Eagly et al., 2000) or instrumentality (task-oriented) and expressiveness (socioemotional) (Spence & Buckner, 2000; Spence, Helmreich, & Stapp, 1974). Agentic classifications are associated with assertion and control. Traits associated with agentic beliefs are aggressive, ambitious, dominant, self-confident, forceful, self-reliant, self-sufficient, and individualistic (Eagly & Sczesny, 2009). Communal is characterized as compassionate treatment; these traits are commonly affectionate, helpful, friendly, kind, sympathetic, interpersonally sensitive, gentle, and soft-spoken (Eagly & Sczesny, 2009). In connection to gender, men are consistently described in agentic terms and women in communal terms. Along with the Stereotype Content Model and the dichotomy of how women can be stereotyped, traditional women are viewed as being high on communal traits and low on agentic traits, whereas, non-traditional women are thought to possess agentic traits and less communal traits (Cuddy et al.,

2009). The social role theory clearly presents the systematic and theoretical structure of gender stereotypes.

Reflecting on both stereotype models, Stereotype Content and Social Role Theory, there is a connection to personality factors that is important to outline. The stereotype models provide the structure to assess and understand what stereotypes are and the associated trait descriptors. Many of these descriptors or adjectives align with personality factors. This alignment through language makes it able to connect specific social stereotypes to its corresponding personality trait/item. If personality assessments are activating stereotypes by way of the lexical hypothesis, it can explain how score differences by gender are created through the measures.

**Assessing Gender Differences on Personality Assessments**

With a foundational understanding of stereotypes and language along with the context it provides to personality items, the purpose of the study is to examine how personality items function and the source of potential differences between men and women in personality. This research is specifically testing if the item itself is a driver of these personality differences. A methodology that allows for an analysis at the item level is item response theory (IRT) and the assessment of DIF, which determines which items operate differently among of group of individuals.

***Item Response Theory (IRT)***

IRT is a psychometric approach that defines an individual's responses to an assessment as a function of the characteristics of the items and the person's standing on the latent trait (Drasgow & Hulin, 1990; Lord, 1980; Nye, Allemand, Gosling, Potter, & Roberts, 2015). The premise of this statistical approach is that the responses on a measure are indicative of an underlying latent trait. The tenets of IRT state that an endorsement of an item is a result of the

qualities of the individuals and one or more qualities of the item (Furr & Bacharach, 2013). Based on statistical modeling, IRT shows the probability of a person responding to an item in a particular way based on a psychometric profile. In order to use IRT models, two primary assumptions must be met: the underlying construct is unidimensional and that there is local independence, such that when the latent trait is controlled for, there is no relationship between item responses.

There are a variety of IRT models. The models vary based on the number of estimated item parameters and the scoring of the data that is used to estimate the parameters. On many personality assessments, there are no 'correct' responses, the response options are ordered, and the high numbered responses are assumed to represent a higher value on the latent trait. Thus, the appropriate IRT model for those types of assessments is a graded response model. A graded response model operates with the assumption that the individual's value on the latent trait (e.g., personality factor) corresponds with the response option selected by the individual. If an individual chooses the response "very inaccurate", his/her value on the latent trait is expected to be smaller than in an individual that chooses "moderately inaccurate." These models produce option characteristic curves (OCC) (see Figure 1). The OCCs show the latent ability related to the probability of selecting a particular response option. In the model, the number of response options are $m_i$. Both parameters, location and discrimination are estimated for $m_i$ -1 boundary response functions. Boundary responses functions indicate the cumulative probability of selecting a response option equal to or higher than the current response option. After the parameters are estimated, the next step is to equate the two groups.

Equating is the process of adjusting for differences in groups so that scores can be meaningfully compared. Here, the item parameter estimates are equated onto a common scale.

This is done by finding a common set of items that contain no DIF. One way to complete this is the mean sigma method. The mean sigma method uses means and standard deviations of the difficulty parameter estimates from the common items to determine *A* and ß coefficients (Marco, 1977). A set of linking constants are then estimated in order to have test characteristics curves from different groups as similar as possible. Once the data is on a common scale, the item response curves (IRC) can be compared across groups to identify DIF. The mean sigma method is summarized through the following formulas. The first equation illustrates that the estimation of the ß coefficient. The means of the category of the common items from both scales are calculated and substituted for the parameters in the equation. The second equation is how the *A* constant is estimated. The standard deviations of the *b* parameter estimates of the common items from both scales are calculated and then substituted for the parameters.

$$ß = \mu(b_I) - A\mu(b_J) \quad (1)$$

$$A = \frac{\sigma(b_J)}{\sigma(b_I)} \quad (2)$$

The logic behind testing for DIF is that after controlling for the latent traits, there should be no relationship between group membership and the response to an item. To compare the item parameters for the two groups (e.g., men and women), several methods are identified. First, chi-square tests are tested for significance (Lord, 1980). Chi-square tests are indices of the difference between the expected frequency of responses for the options to the observed frequency. If the parameters are the same for both groups, the chi-square will not be significant. Second, Raju, van der Linden, and Fleer (1995) differential functioning of items and tests (DFIT) is often used. DFIT is based on the method that if the expected scores are the same for the focal and reference group, the ICC or OCC should not be significantly different. Non-compensatory (NCDIF) examines the potential differential functioning of an item while isolating it from the DIF

information in the other items. This index for an item, is a function of the difference in expected scores for the focal and reference group after accounting for the latent trait. Significant differences mean that one of the groups are endorsing the item differently. NCDIF is similar to Lord's chi square analysis. Large NCDIF index values are present if there are large differences between men and women exist when the latent trait is controlled for. The suggested cutoff value for an item to possess NCDIF is .096 (Raju, 2000). To classify that significant DIF exists, there must be a significant chi-square and a NCDIF greater than .096. From these analyses, an effect size indicates the magnitude of differential item functioning.

Exploring personality differences focused on the content of the item requires a direct look at the item level. Moreover, the use of IRT is an addition to the personality literature. IRT and the measurement of differential item functioning (DIF) supports the interest in investigating if it's possible to predict which items may function differentially based on stereotype coding.

**Overview of Hypotheses**

The crux of this research is that personality is inherently linked to stereotypes through language. With this connection, it is plausible that stereotypes are embedded in how personality tests are constructed and thus how personality is measured. Stereotypes are prevalent social influences, and oftentimes, act as markers of how individuals should behave. Since stereotypes are guiding forces, when they are present, they are critical pieces of information. Stereotypes communicate behavioral expectations thus influencing the way individuals act or the beliefs to which they subscribe (Eccles, 1993; Mitchelson et al., 2009). This impact is also expected to appear in personality measurement based on the communality of language between both personality and stereotypes. The presence of the stereotype acts as an influence that may impact

the expected endorsement by men and women that might otherwise demonstrate identical scores on the latent personality trait.

Two studies were conducted to examine the role of stereotypes on the observed gender differences on personality assessment scores. The first study examined if the occurrence of differential item functioning (DIF) is associated with the presence of gender stereotypical content embedded in the items. The Stereotype Content Model and the Social Role theory will be used to hypothesize which items are expected to function differently by gender.

In the second study, an experimental method was employed whereby the gender stereotypical content embedded in the items was manipulated to reflect opposite ends of the gender stereotype continuum. The effects of this manipulation on the personality scores of men and women was examined. It was expected that the stereotype content in an item will interact with the respondent's gender such that scores will be lower in the conditions where the stereotype content is the opposite of what is traditionally associated to the participant's gender. Both studies are described in the following sections.

## Study 1

This first study looked to predict which personality items would show DIF. Here, coders assessed and rated archival personality data based on the Stereotype Content Model (Fiske et al., 2002) and Eagly's (1987) Social Role Theory. Some gender differences may be artifactual and thus driving DIF. Items were coded and classified on stereotypes dimensions of competence, warmth, agentic, communal, or neutral (the absence of stereotype content). After coding, the study looked if the presence of DIF is more likely to be observed in the items that possess stereotypic content. Overall, the two goals of this study were to first assess the presence of

gender-specific DIF on personality items and secondly, to understand and predict which items will show DIF using a theoretical basis. The hypotheses for Study 1 were:

Hypothesis 1: Personality items with gender stereotypes embedded in them will show a higher frequency of DIF than items without gender stereotypes embedded in them.

Hypothesis 2: Personality items that contain agentic content will show a higher frequency of DIF in the direction of men compared to communal or neutral items.

Hypothesis 3: Personality items that contain communal content will show a higher frequency of DIF in the direction of women compared to agentic or neutral items.

Hypothesis 4: Personality items that contain warm content will show a higher frequency of DIF in the direction of women compared to competent or neutral items.

Hypothesis 5: Personality items that contain competent content will show a higher frequency of DIF in the direction of men compared to warm or neutral items.

## Method

### Data

The archival data used in study 1 comes from Scherbaum (2003). From this research, the control condition sample was used. A total of 560 participants were in this condition. 274 participants reported their gender and were used for this analysis. Sixty percent of the sample identified as female.

### Measures

#### *International Personality Item Pool*

The International Personality Item Pool (IPIP; Goldberg, 1999) is a measure of the five-factor model of personality. The IPIP consolidated personality-descriptive items and provided an instrument for the scientific community (Goldberg, 1999; Hendriks, Hofstee & de Raad, 2002).

It is similar to the NEO-PI-R (Costa & McCrae, 1992) and measures conscientiousness, extroversion, neuroticism, agreeableness, and openness. The IPIP version that was used was the NEO 30 Factor.

The IPIP/NEO 30 is comprised of 5 scales, one for each of the five personality factors. Each scale has six sub-scales with 10 item that measure the subfactors of the five higher-order factors. All of the five-factors have a total of 60 items, resulting in the full IPIP/NEO 30 being 300 items. Instructions to the IPIP asks respondents to select the response option that most describes them. Items are on a 5-point scale from 1 to 5, with 1 representing "very inaccurate" to 5 representing "very accurate." All items are presented in Appendix A.

Reflecting on the IPIP/NEO 30, evidence has suggested the inventory has satisfactory psychometric properties. Scores on the IPIP have been found to have high internal consistency ($\alpha$ = .80) (Goldberg, 1999). There is also evidence for construct validity and that the measure does indeed assess only five factors, with all items loading correctly on the appropriate factor (Goldberg, 1999).

***Personality Item Coding***

All 300 IPIP personality items in the dataset were coded as to the level they reflect agentic, communal, warm, and competence stereotypes. Six trained coders (three pairs of coders, one male/ one female) coded each item on rating scales that reflected the level of each stereotype factor that is perceived in the item. Each pair of coders coded all 300 items on one set of rating scales, either the warmth/competent scales, agentic/communal scales, or the masculine/feminine scales. The masculine and feminine scales were included for exploratory analysis (see Appendix B for the coding sheet and Appendix C for the training guide). The consistency of ratings was assessed with an interrater reliability index using an intra-class correlation coefficient (ICC).

This allowed for an understanding of the agreement on the classification of these items on the assigned scale.

Items were considered possessing the stereotype (agentic, communal, warm, competent, masculine, feminine) if the average rating was 2.0 or higher. This cutoff was determined as on a three-point scale it represents a rating of at least somewhat by both raters. Items with an average rating of 1.0 on both scales was classified as neutral. The level of agreement was assessed using intraclass correlations. Disagreements were addressed through active discussions to determine coding decisions.

## Results

### Coding Analysis

Intraclass correlations were computed to measure the level of agreement of the ratings/codes on each stereotype dimension. ICC estimates were based on a mean rating (k=2), consistent agreement, two-way mixed effects model. Initial ICCs are reported as follows: codes on the warmth dimension resulted in moderate reliability (.699). Codes for the competent dimension resulted in moderate reliability (.771). Codes for the agentic dimension resulted in moderate reliability (.617). Codes for the communal dimension resulted in moderate reliability (.564). With the exploratory analysis of the masculine and feminine dimensions, codes for the masculine dimension resulted in poor reliability (.460) and codes for the feminine dimension resulted in moderate reliability (.506). Following discussions to resolve disagreements, the final ICCs were as follows: agentic codes (.765), communal codes (.761), warm codes (.768), competent codes (.850), masculine codes (.703), feminine codes (.753).

Assessing the average rating of each item, approximately 19% of the items were rated as highly representative of the warmth stereotype dimension and approximately 20% of the items

were coded as highly representative of the competent stereotype dimension. The items that were coded on the warmth stereotype were present across all of the five factors, with the majority on extraversion (30%), agreeableness (25%), and openness (22%). Items that were coded as possessing the competence stereotype were across all five factors and most prominently on the conscientiousness factor (40%). Looking at the agentic stereotype dimension, approximately 23% of the items were coded as highly representative of this dimension. The items found to possess agentic stereotypes were present in all five factors. The majority of those items were on the conscientiousness (37%) factor, followed by extraversion (19%) and neuroticism (19%) factors. The communal dimension was rated as highly present in approximately 9% of the items. These items were represented across all five factors with the greater representation on the agreeableness (45%) and extraversion (41%) factors. With the exploratory assessment of masculine stereotype presence in the items, 6% of the items were coded as highly representative of masculinity. Similar to other items, there was a spread across all personality factors with most on the conscientiousness (30%) factor, followed by extraversion (25%). Lastly, around 6% of the items were coded as reflecting feminine qualities. These items were found across all the five traits, with the majority present on the agreeableness (33%), extraversion (22%), and openness (22%) factors. These reflect the final percentages/decisions based on the final coding agreements. Overall, there was a modest representation of items that was coded to possess stereotype content and they were reflected across all the personality traits in varying amounts.

**Descriptive Statistics**

Descriptive statistics were computed for the IPIP items. Means and variances for each item are reported in Table 4. The internal consistency reliability estimate of the neuroticism scale was 0.853; the internal consistency reliability estimate of the extraversion scale was 0.836; the

internal consistency reliability estimate of the openness scale was 0.740; the internal consistency reliability of the agreeableness was 0.788, and the internal consistency reliability estimate of the conscientiousness scale was 0.841.

**Tests for Gender Differences on Personality Items**

***Score Level Differences***

To determine if the gender of the participants impacted the responses to the personality items, a MANOVA was performed. Gender was entered as the independent variable and the five personality factor scores as the dependent variables. There was no significant multivariate effect, Pillai's trace= 0.027, $F(5, 268) = 1.491$, $p < .07$, partial $\eta^2 = 0.027$. The univariate $F$ tests revealed that the men and women were significantly different on Neuroticism, $M_M = 161.46$, $M_F = 170.22$, $F(1, 5032) = 5.139$, $p < .024$, partial $\eta^2 = .019$. No significant differences were found for the other personality factors (Conscientiousness, $M_M = 211.22$, $M_F = 210.62$; Extroversion, $M_M = 201.94$, $M_F = 198.22$; Openness, $M_M = 202.50$, $M_F = 204.17$; Agreeableness, $M_M = 200.82$, $M_F = 202.85$).

***Item Level Differences***

Looking at specific items, there were significant differences on particular items between men and women. It was found that 33 of the 300 items showed a statistically significant difference between gender mean scores on the item. Table 5 presents the mean item scores and partial eta squared for each of the 300 items. Each personality factor has at least one item that had significant differences in scores between men and women. The neuroticism factor presented the most items in which men and women scored differently.

***Factor Analysis***

A principal axis factor analysis was conducted to determine if the unidimensionality assumption was met before IRT analysis. This assumption indicates that there is one factor

underlying the responses and the first factor should account for 20% of the variance (Reckase, 1979). This analysis was performed on each factor. A promax rotation was chosen to account for the five factors not being completely orthogonal (Costa & McCrae, 1992) and to examine the stability of the factor structure. Table 6 presents the first and second eigenvalue from the PFA using the subfactor scores and the percent of variance explained. The first component from the PFA analysis for extraversion and neuroticism accounted for more than 20% of the variance. Conscientiousness accounted for slightly less than 20%. Openness and agreeableness accounted for less than 20% of the variance. All five traits had a dominant first factor.

In order to determine how many factors were present, a parallel analysis was used. Parallel analysis is procedure in which eigenvalues random data matrices are computed based on the number of observation and variables in the original data (Horn, 1965). Then, the eigenvalues estimated from the original or actual data are compared to the eigenvalues from the random dataset. If the first eigenvalue is higher than the eigenvalue computed from the random dataset and the second eigenvalues is lower than the random eigenvalue, the unidimensionality assumption is satisfied. The parallel analysis was performed using a parallel analysis engine (Patil, Singh, Mishra, & Donavan, 2017). Table 7 shows that there are five factors, as the five eigenvalues are larger than the eigenvalues from the random dataset. This presence of five factors also indicates that the unidimensionality criterion was not met.

Regarding unidimensionality, if the strict assumption is not met, there is still a question of how much can the assumption be violated to still produce stable parameter estimates for IRT analysis. When determining next steps, there is research that suggests if a dominant first factor is present, IRT models will estimate that first factor (Reckase, 1979). Additionally, graded response IRT models are somewhat robust to violations of the unidimensionality assumption (Kirisci, Hsu,

& Yu, 2001). Although this assumption is violated, due to the presence of a clear first factor, the parameter estimation is unlikely to be impacted.

Looking at the loading on the present five factors, the rotated factor structure matrix mostly loaded on a five-factor model with overlap on some subscales (Table 8). Many of the subscales of conscientiousness loaded on the correct factor, except first subscale (self-esteem) and fourth (achievement striving) and fifth (self-discipline) subscale. These three subscales appeared to load on factor for extraversion. One additional area of cross loadings was with the openness subscales. The first and third subscales loaded on the factor of neuroticism and the fourth subscale loaded on extraversion.

***Parameter Estimation and Model Fit***

The XCalibre program (version 4; Guyer & Thompson, 2014) was used to compute the marginal maximum likelihood estimates of the item parameters and the expected a posteriori estimates of the theta parameters in Samejima's (1969) graded response model. The parameters were estimated for each of the five personality factors separately. Thus, five sets of item and person parameters were estimated for each IRT model. The program defaults were used in all analyses. The results of the IRT analyses for the polytomously scored items for each individual and each item are available upon request.

In order to establish model fit to the data, MODFIT using chi-square statistics recommended by Drasgow, Levine, Tsien, Williams, and Mead (1995) was employed. This methodology indicates the differences between the expected frequency of the responses for the options and the observed frequency of responses. MODFIT computes fit plots and chi-squares for items singles, doubles, and triples (Drasgow et al., 1995). When interpreting this analysis, a good fitting model has a chi-square to degree of freedom ratio less than 3 for singlets, doublets,

and triplets. The chi-square results for all five factors are presented in Table 9. The results suggest good model fit, as $\chi^2$ were less than 3.

***IRT Model Overview***

This study used Samejima's (1969) graded response model to estimate the parameters for both men and women. As the IPIP has five response options, four boundary response functions were estimated. From the pattern in the archival data, the strongly disagree and disagree response options were collapsed. For this study, the equating method that was used is the mean sigma method. The non-compensatory DIF (NCDIF) index was used.

**Differential Item Functioning.** After assessing the parameters and fit, differential item functioning (DIF) analysis were performed. The first procedure was the Mantel procedure (1963) which examines the relationship between two variables in a 2 x *K* frequency table (*K* = number of response options). This procedure controls for the level of a third variable. The relationship is measured as an odds ratio and the degree of DIF. The null hypothesis is that the odds ratio is 1.0. An odds ratio greater than 1.0 indicates that the group coded as '1' has higher odds of endorsement, even after controlling for the level of the latent trait. An odds ratio less than 1.0 indicate that the group coded '1' has lower odds of endorsement. Significant values are interpreted as evidence of DIF, such that individuals in the focal and reference group differ on their endorsement of an item, even after controlling for the latent trait.

To conduct this procedure, the groups were coded; females were coded as '0' for the reference group and males were coded as '1" for the focal group. A 2 x 4 frequency table was used as there were four response options. The four response options are result of collapsing, or the least endorsed option of 'strongly disagree' with 'disagree'. The stratification variable in each analysis was the composite personality factor score for each of the five factors (e.g., sum of

the neuroticism items, sum of the extraversion items, sum of the agreeableness items, sum of the openness items, sum of the conscientiousness items). In order to ensure that there were adequate sample sizes at each level of the stratification variable, the stratification variable (e.g., the total score) was collapsed into six categories of personality factor scores and these categories were used.

The Mantel procedure designates an index number to each ordered response option. Using this designation, the procedure then compares the item means for the focal and reference groups that are matched on the stratification variable. Applying this to the data, women and men matched on the same level of the personality factor were compared by item means to identify DIF. Table 10 indicates the results of the Mantel procedure and the items reflecting differential item functioning. Approximately 9% (26 of 300 items) of the items indicated that there was DIF present. Each personality factor had at least one item that function differently. The trait with the greatest number of items that were differential functioning was agreeableness (7 items), followed by conscientiousness (6 items) and openness (6 items). The trait with the least number of items was extraversion (3 items).

Additionally, the latent trait approach of IRT was run. This analysis requires that the item parameters were equated on a common metric using the mean-sigma method (Marco, 1977). As aforementioned, the mean-sigma method uses the means and standard deviations from each group to determine the slope and intercept coefficients for the linking equation to equate the two groups. Specifically, with a graded response model, the means and standard deviations are used instead of the item difficulty parameters in the 1-PL, 2-PL, and 3-PL model (Cohen & Kim, 1998). Table 11 shows the linking coefficients from the mean-sigma method used for the latent DFIT analysis.

The IRT DIF method used is an adaptation of Raju et al.'s (1995) DFIT through a newer R package also named DFIT created by Cervantes (2017). The DFIT R package assess differential item functioning and item parameter replication. The particular index of focus is non-compensatory DIF (NCDIF). NCDIF measures the probability of endorsing a response between two different groups while assuming that all items are free of DIF. NCDIF index values are large when there are major differences between the focal and reference (i.e., women and men) groups expected scores after controlling for the latent trait. To determine significance of NCDIF, cutoff values have been established based on simulations. The recommended cutoff for polytomous items with three, four, five response options is 0.096 (Kleinman & Teresi, 2016).

Findings from the DFIT analysis are presented in Table 12. Seven items out of 300 items were identified as differential item functioning based on the cutoff of NCDIF. Three of the items were on the neuroticism trait (Figure 3, 4, 5), two were on the agreeableness trait (Figure 6, 7), and two were on the openness trait (Figure 8, 9). There is a noticeable difference in the amount of DIF items through the observed score and latent trait methods. This difference may be attributed to the conservative standard and complexity of the NCDIF index with polytomous data. To test the hypotheses, items that were classified as DIF through both approaches, the Mantel procedure and latent trait approach, were used.

***Hypothesis Testing***

Hypothesis 1 (H1) stated that personality items with gender stereotypes embedded in them would show a higher frequency of DIF than items without gender stereotypes embedded in them. Approximately 70% of the items that were found to shown DIF were coded to possess one of the stereotypes (agentic, warmth, competent, masculine, or feminine). A majority of the items

that were found to function differently between men and women were identified to include stereotype language. Thus, this hypothesis was supported.

The following hypotheses focused on the pattern of DIF found in the items. H2 hypothesized that items that contained agentic stereotype content would show a higher frequency of DIF in the direction of men compared to communal or neutral items. Of the two items that were coded as agentic, DIF was found in the direction of men compared to communal or neutral items. These results support this hypothesis. Opposite of the agentic hypothesis, H3 stated that items that contain communal content will show a higher frequency of DIF in the direction of women compared to competent or neutral items. However, none of the items that shown to have DIF were coded as having communal content. With the absence of communal content found in the identified DIF items, the hypothesis was not supported.

Regarding the warm and competent items, H4 hypothesized that personality items that possessed warm content would show a higher frequency of DIF in the direction of women compared to competent or neutral items. In contrast of the expectation, all of these items were found to be more likely endorsed by men than women. This hypothesis was not supported. With the competent content, H5 expected that personality items that contained competent content will show a higher frequency of DIF in the direction of men compared to warm or neutral items. 85% of the competent content items that possessed DIF were more likely to be endorsed by men compared to warm or neutral items, supporting the hypothesis.

Lastly, looking at the exploratory stereotype content of masculine and feminine stereotypes, only two of the items that showed DIF were coded masculine/feminine. This small portion of items does not provide the opportunity to find an established pattern.

**Study 1 Discussion**

The purpose of Study 1 was to predict the presence of DIF on the IPIP through the identification of stereotypic traits in the item content. Overall, there was partial support for the hypotheses presented in this study. Approximately 13% of the items were coded as reflecting the presence of either warmth, competent, agentic, communal, masculine, or feminine traits. Looking at the items for the presence of DIF, approximately 9% of the items were found to function differently between men and women. It is was not expected that a high number of items would show DIF. This finding is in line with previous research where scales may have shown DIF, there can still be a minimal number of individual items that possess DIF (Mitcheson et al., 2009). Within these items, there was alignment with stereotypes that were associated with men (agentic, competence) showed DIF in the direction of men. Only one item ("know the answers to many questions") exhibited DIF in the direction of women.

The study is one of few within the personality literature that has assessed the content of personality items for stereotypic content in order to predict if the item will function differently between men and women. The results of this study indicate that a moderate portion of items were viewed as possessing stereotypic content. Connected to the presence of DIF, 70% of the items that were found to have a presence of stereotypic traits also functioned differently between women and men. In line with the expectation, the presence of stereotypes provides an explanation of the level of endorsement by gender. While the total amount of items that showed DIF was minimal, a sizeable number of items that had DIF were also coded as possessing stereotype-related content. This finding is insightful as it appears that the wording/content of the item can lend an indication of the item may function differentially and that it may be possible to predict which items may possess DIF in advance. For psychometricians, the potential to predict can support further personality scale construction and revision.

## Study 2

The purpose of the second study was to investigate the impact of stereotype content in items and to causally test that stereotypes embedded in the items impact individuals' response behavior. The study examined personality items through an experimental lens by rewording select IPIP items to create items that reflect neutral and stereotyped content in order to determine the impact of the addition or removal of stereotypes on item responses. It was expected that exacerbating the stereotypes or elements of stereotypes would result in interactions between stereotype, item content, and gender. In other words, the same personality item was manipulated to reflect differing levels of stereotype content. This effect would illustrate the interconnectivity of language that binds stereotypes and personality. The hypotheses for study 2 are:

Hypothesis 6: There is an expected interaction between agentic/communal item content and gender such that men will have higher scores on agentic items, whereas women will have higher scores on the communal items.

Hypothesis 7: There is an expected interaction between warm/not warm item content and gender such that women have higher scores on the warm items, whereas men will have higher scores on the not warm items.

Hypothesis 8: There is an expected interaction between competent/not competent item content and gender such that men have higher scores on the competent items, whereas women will have higher scores on the not competent items.

### Pilot Study

A pilot study was conducted to test the manipulation of the items and if they reflect the stereotype they were intended to represent. Items that were coded as possessing stereotype content and showed DIF and/or significant mean differences between men and women in Study 1

were selected for the pilot. These items were reworded to make the item more neutral with the purpose of removing the stereotype content that was coded to be present in the item. Additionally, the original item was rewritten in the contrasting version to see if the stereotype can activated be in the opposite direction. This resulted in four sets of items (agentic/neutral/communal, warm/neutral/not warm, competent/neutral/not competent, masculine/neutral/feminine) (Appendix D). Similar to the first study, the masculine/feminine items were included for exploratory analyses. To test the strength of the manipulation, the original, neutral, and rewritten item were rated on each respective scale and assessed for agreement.

***Pilot Study Participants***

Each set of items was rated by a group of participants recruited from Amazon MTurk and the Baruch College Psychology Department's participant pool. When testing the agentic/communal items, the pilot group included 16 participants recruited from MTurk. Approximately 68% of the sample self-identified as male and 32% self-identified as female. Lastly, the warm/competent participants were recruited from the Baruch College Psychology Department participant pool. The pilot participants ($n$ = 13) were evenly split with 50% self-identifying as male and 50% self-identifying as female. For the masculine/feminine items, the pilot group consisted of 19 participants recruited from Amazon MTurk. Similar to the other pilot study, 68% of the sample self-identified as male and 32% self-identified as female.

***Pilot Study Procedure***

Participants received one of the three measures to rate on the associated scale (Appendix C). Participants that received the warm/competent measure rated the warm-based items as either warm, neutral, not warm and the competent-based items as either competent, neutral, not

competent. With this scale, participants were asked to select the rating that best described the personality item they were reading. An example of a warm item and not warm item is “Rarely complain/ Often complain.” An example of a competent and not competent item is “Know the answers to many questions/ Often don’t know the answers to many questions.” This same procedure was followed with the agentic/communal and masculine/feminine measures. Participants given the agentic/communal items rated the items as either agentic, neutral, or communal. Those given the masculine/feminine measure rated the items as either masculine, neutral, or feminine. One example of agentic/communal items is “Value competition over cooperation”/ “Value cooperation over competition”. An example of masculine/feminine items is “Don’t notice my emotions/ Experience my emotions intensely."

***Pilot Study Results***

Regarding the agentic/neutral/communal items, 85% of the items that were originally coded or rewritten to reflect the agentic stereotype were rated as agentic. 71% of the items that were rewritten to be neutral were rated as such. All of the items (100%) that were originally coded or rewritten to reflect the communal stereotype were rated as communal by the participants. The agreement among the raters was high (.917). These items appeared to reflect the stereotypes that were expected. Disagreements were addressed by calculating the mean of all the ratings to determine the agreed upon rating.

With the warm items, 100% of items that were originally coded as warm were rated as warm by the pilot group. 60% of the items that were rewritten to be neutral were perceived as reflecting a neutral item. Lastly, all of the items (100%) that were rewritten to reflect the opposite of warm were rated as intended. There was strong rater agreement (.920) among these items. Overall, the items in this category operated as expected.

All of the items (100%) that were originally coded as competent were also rated as competent by the pilot sample. 60% of the items that were rewritten to be neutral were rated as neutral. Lastly, all of the items (100%) that were written to represent "not competent" were rated as intended. Rater agreement was strong (.956) across all items.

For masculine/neutral/feminine, 88% items that were originally coded as feminine or rewritten to reflect the feminine stereotype were rated as feminine on the rating scale. 38% of the items that were rewritten to be neutral were rated as neutral. These items were reviewed and reassessed. 100% of the items that were originally coded or rewritten to reflect the masculine stereotype were rated as masculine. There was high agreement among raters across the items (.974). Thus, most of the items were found to operate as intended. The pilot study provided support that the items reflected the stereotypes they were manipulated to represent, thus allowing it to be the items used in Study 2.

**Main Study Participants**

The sample included approximately 212 participants. Participants were recruited from the Baruch College Psychology Department's participant pool. The sample comprised of approximately equal self-identified male (55%) and female (44%). Approximately 92% of the sample reported their age between 18 and 24. 45% of the participants identified as Asian, 26% of identified as Hispanic, 20% identified as White, 8% identified as Black, with 6% identifying as Other or prefer not to answer. 65% of the sample reported completing a personality survey/questionnaire in the past. Close to 17% reported that have taken a personality survey in the past but were uncertain.

**Design**

The design of the experiment was a 2 (item type: original item, re-written, opposite item) x 2 (gender: male, female) mixed design. The within-subject factor was item type and the between-subject factor was gender. For the within-subjects factor, participants completed items across all the stereotype categories (agentic, communal, warm, competent, masculine, and feminine). Participants were given all of the original personality items, the neutral item versions, and the rewritten items which reflected the opposite stereotype from the original.

**Measures**

***Manipulated Personality Items***

Original items from Study 1 were manipulated to include the original item, neutral-written item, and the opposite direction stereotype-written item. The inclusion criteria for the items included in Study 2 was the items function differentially and/or showed mean differences by gender in Study 1. A 66-item version was administered to participants (12 agentic/neutral/communal items, 21 warm/neutral/not-warm items, 21 competent/neutral/non-competent, 12 masculine/neutral/feminine items (Appendix D). One example of the agentic/neutral/communal items is “Remain calm under pressure”/ “Remain at ease under pressure”/ “Help others stay calm under pressure”. Examples of the warm/neutral/not-warm items are “Radiate joy”/ “Occasionally joyous”/ “Don’t often feel joy”. An example of the competent/neutral/non-competent items are “Excel in what I do”/ “Excel in some things that I do”/ “Don’t excel at anything I do”. Lastly, examples of the masculine/neutral/feminine items are “Don’t understand people who get emotional”/ Don’t understand peoples who are affected by their feelings”/ Sympathize with those who get emotional”. Given the inclusion criteria of the study, there were no complete personality scales administered. The order of the 66 items on the measure were randomized for each participant.

### *Demographic Characteristics*

Participants were asked to report demographic characteristics such as sex, age, and race. Participants also reported if they had experience with personality inventories.

### Procedure

Participants volunteered to complete a personality inventory and started the study by completing the informed consent procedure. Participants read the instructions and were given the full 66-item version of the personality inventory. After completing the personality items, participants answered demographic items and were then thanked for their participation.

## Results

Descriptive statistics were computed for all modified and original IPIP items. To test the hypotheses, a two-way mixed ANOVA was conducted to detect mean differences between men and women on the original and opposite stereotype-rewritten version of the item.

### Manipulation Check

The selection of items used in Study 2 were based on the findings in Study 1. As the item level was of interest, t-tests were conducted on all items in each condition (agentic/communal, warm/ not warm, competent/ not competent, masculine/feminine) to determine if item endorsement by men and women changed based on the wording of the item. Table 13 reports the means, standard deviations, and significance values. Overall, the pattern of endorsement of the original IPIP items from Study 1 (i.e., items that were not manipulated) was replicated in Study 2 (e.g., items that were primarily endorsed by women in Study 1 were also endorsed by women in Study 2). This suggests that the gender preferences for the items are potentially consistent across different student samples.

As expected, the neutral items showed no significant differences between men and women. Thus, there is initial evidence that the removal of the stereotype also removed the gendered preference for the items. Moreover, the revisions to the original items to reverse the stereotype reflect in the item appeared to impact participant response behavior. In approximately 60% of the items, the direction of the endorsement completely reversed from the original item. For example, the original, masculine-coded item, "am not easily affected by my emotions" was endorsed by men (*M*=2.91) more than women (*M*=2.73); the re-written, feminine item "am easily affected by my emotions" was endorsed by women (*M*=3.45) more than men (*M*=3.03). Taken together, these data provide initial evidence of the effectiveness of the manipulation.

**Item Type Analysis**

Along with the comparison of all the item types, the original IPIP item, the re-written neutral item, and the re-written opposite stereotype, it is of interest to explore if the manipulation of the item content from the original stereotype (e.g., agentic) to the opposite stereotype (e.g., communal) effectively reversed the pattern of endorsement on the original item. The mixed ANOVAs were ran comparing individual items as the wording and content change was the level of interest. Results are categorized by the stereotype code of the original item. Table 14 presents the effect sizes of each main effect and interaction.

***Agentic/Communal***

In H6, it was hypothesized that men will have higher scores than women on newly written agentic items and women will have higher scores than men on newly written communal items. As can be seen in Table 14, most of the items in this communal/agentic category showed minimal significant effects with the manipulation of the item and the interaction with gender.

Only 25% of the items showed results that supported the hypothesis. Thus, this hypothesis was not supported.

***Warm/Not Warm***

In H7, it was hypothesized that men will have higher scores than women on warm items and women will have higher scores than men on newly written not warm items. As can be seen in Table 14, the items in the warm/not warm category presented some unexpected trends when significant interactions appeared. Approximately 29% of the items produced a significant interaction between gender and item type. Of the 4 items showing significant interactions, the results were in the opposite direction of the hypothesis. Thus, this hypothesis was not supported.

***Competent/Not Competent***

In H8, it was hypothesized that men will have higher scores than women on competent items and women will have higher scores than men on not competent items. As can be seen in Table 14, the competent/non-competent items showed variation in endorsement by gender. Two showed a significant interaction. These interactions showed significance in the opposite of the hypothesized direction. Women were more likely to endorse the original, competent-coded item than men. The remaining items (70%) did not show any significant results. In summary, the hypothesis was not supported.

***Masculine/Feminine***

The masculine and feminine items were included in this study as an exploratory analysis of stereotype— based language. The expected interaction between gender and item type such that men were more likely than women to endorse masculine items and women were more likely than men to endorse feminine items. All of the item pairs in this category supported the direction of the hypothesis as presented in Table 14. Each pair of items exhibited a significant interaction

between item type and gender. When the item was coded to be masculine, men were more likely to endorse the item. As the item was rewritten to be more feminine, women became the prominent endorser. Endorsement by gender aligned with the associated stereotype. Furthermore, the manipulation of the item effectively activated the opposite stereotype. Masculine and feminine stereotypes resonated with participants, whereas they were impactful and guided the endorsement.

## Study 2 Discussion

The purpose of the second study was to take a more experimental approach to understand the relationship between personality items, stereotypes, and gender. Overall, there were mixed results that the language used in personality items can influence the level of the content and the stereotypes embedded in them. As the items were manipulated into two versions of the original item, neutral and the opposite stereotype, each change moved the endorsement in the intended direction. Nearly all rewritten neutral items, outside of one, showed no significant differences between genders. This occurred even when the original item exhibited a significant difference between men and women. Moreover, some of the items that were re-written to showcase the opposite stereotype reflected a significant interaction.

When looking at the relationship between gender and item type, the items that were coded as masculine and feminine resulted in the most significant interactions, whereas, the items reflecting agentic, communal, competent, and warm stereotypes showed few statistically significant interactions. One explanation for this pattern of findings may be that stereotypes such as agentic and communal have remained prominent in public opinion and social roles (Eagly, Nater, Miller, Kaufmann, & Sczesny, 2020). In the Eagly meta-analysis, 97% of respondents stated that women are more communal in a 2018 poll, which was a 14% increase from the same

poll administered in 1983. Additionally, agency was consistently rated as a descriptive of men over time. Seeing that these stereotypes have remained strongly associated with the assigned gender, the manipulation of the items to reflect agentic/communal may have resonated with participants. It is also lends support to the strong presence of these stereotypes in language, even today. Regarding the exploratory analysis of masculine/feminine, these stereotypes are similar in theme and content to the agentic/communal dichotomy.

The warm/not warm, competent/not competent items did not show as a prominent of an impact on gender. Most of the interactions were non-significant. For the few that did produce significant interactions, the pattern was opposite of hypothesized. Men were found to endorse warm items and women found to endorse the not-warm items. Also, women were found to endorse competent items and men were found to endorse the not-competent items.

The Eagly et al. meta-analysis also provides insight into the competence stereotype. They found that the competence stereotype has changed over time as 34% of respondents found women more competent than men in 1946 and, in 2018, 65% found women more competent than men (Eagly et al., 2020). Additionally, agency is found to be a stronger theme than competence for male stereotypes (Sczesny, Nater, & Eagly, 2019). The trends in this study and the overall meta-analysis provide some explanation on the strength and presence of the stereotypes in these rewritten items.

Regarding both the warm and competent items, there was an established pattern by item type. The original item was significantly endorsed more often than the re-written item. One possible explanation is that the re-written items may have not resonated with the participants. This would also explain the prevalence of non-significant interactions within these stereotype manipulations.

Transitioning to personality as a construct, the original items were reflective of personality traits. Items that were selected from Study 1 to be included in Study 2 were representative of all five personality traits. Looking at the items that showed significant interaction, four of the five personality traits (agreeableness, conscientiousness, openness, and neuroticism) were reflected. An interesting finding here is that stereotypes can be present across many of the big five personality traits. When assessing personality item content for its language, all items should be reviewed for the possibility of stereotyped content. The manipulated extraversion items did not show a significant interaction of gender differences. Previous research suggests that this trait does not often reflect differences between men and women (Costa, Terracciano, & McCrae, 2001). More specifically, both extraversion items were on the subtrait of cheerfulness. This subtrait is also found to show no differences by gender (Ruch, 2014). As these items reflect a trait and subtrait that historically men and women endorse equally, it is plausible that no difference would be present here.

Overall, this study presents findings regarding stereotype content in personality items and the ability to causally test if the item content can be manipulated to reflect neutral and stereotype content. It was found that it may be possible to remove semblances of stereotypes in the original item as well as exacerbate the opposite stereotype. Many items that originally showed differences by gender that were rewritten to be neutral showed no significant differences afterwards. While not robust, some item pairs resulted in a significant interaction, such that items that were manipulated to reflect the opposite stereotype were endorsed by the group associated with the stereotype. The results of this study offer a foundation to the understanding of the connection between language, stereotypes, and the *content* of personality items.

## General Discussion

Personality inventories have been used for the purpose of explaining and predicting attitudes, behaviors, performance, and outcomes in organizations (Ones, Dilchert, Viswesvaran, & Judge, 2007). These inventories are remarked for their criterion-related validity for job performance. These assessments are also found to be less problematic as they don't often lead to results that are impacted by extraneous variables. However, there has been some evidence of personality assessments showing systematic differences based on gender within the measure. This has led researchers to investigate if these differences are valid psychological differences (e.g., biological, social, evolutionary) or if the differences exist as a result of properties of the personality assessment (e.g., measurement) (Reise, Smith, & Furr, 2001). Moreover, in this examination, the measurement explanation is not one that has been has fully explored. To address this gap, the purpose of this research was to look at the construction and language of personality items using theory to predict and ascertain if the item content is reflective of stereotypes that ultimately impact the endorsement of the item by gender.

The two studies sought to answer specific research questions. The first study looked to answer if certain personality items function differently between men and women. Furthermore, it looked to answer if it is possible to predict which items may show DIF by assessing the item for stereotype content. The second study focused on understanding personality items and the extent of stereotype impact through an experimental lens by examining the malleability of personality items. The major questions of interest were if personality items can be manipulated to reflect stereotypes and if this would impact the endorsement of the item by gender. This discussion will look to summarize the findings of this research and focus on implications and future directions.

### Review and Interpretation of Results

A focal point of this research was to predict and identify if personality items on the IPIP function differently between men and women. Looking at the findings from a holistic view first, there was a small proportion of items that exhibited DIF. Most items functioned equally between genders. Whereas there was no expectation of a certain number of items, the presence of DIF is aligned with the current research that found DIF within subareas or subtraits of personality assessments (Escorial & Navas, 2007; Sharp et al., 2014; Wetzel et al., 2013). In line with other studies, the presence of DIF was not across the full measure (e.g., Escorial & Navas, 2007; Forrest, Lewis, & Shevlin, 2000; Karanci, Dirik, & Yorulmaz, 2007).

Looking into the DIF that was present, many of those items were coded and identified as possessing gender-related stereotype content. The language used in personality items is impactful to its measurement ability. Identifying items that reflect stereotype language appeared to provide insight into the items that are more likely to lead to the items producing different endorsement between men and women. This finding supports the premise that it may be possible to predict which items may function differently by gender. Most importantly, this finding presents one potential explanation that can explain why differential functioning occurs and that the difference found may not be an inherent difference based on gender, but measurement error.

This study is one of the few to posit predictive hypothesis based on psychometric properties and psychological theory to assess which personality items may function differentially by gender. Findings in this study provide alignment to newer studies that found have also explored DIF in personality measurement. These results indicate that we can review items for indicators of stereotypes and potentially anticipate the presence of DIF when developing personality assessments. Using this awareness, we can critically assess personality items with an eye towards gender-based language.

An additional relationship of interest, specifically in Study 2, was the pattern of endorsement based on the rewritten stereotype present in the item, such that the endorsement would align with the gender-associated stereotype found in the item (e.g., men-agentic, competent stereotype-coded items; women- communal, warm stereotype-items). This pattern was somewhat supported with the manipulated agentic items. Those items were endorsed more by men and the manipulated communal items were endorsed more by women. These results are in line with the stereotype literature. Agency stereotypes are continually associated with men even as a change in gender roles have been explored (Eagly et al., 2020). Through the exploratory analysis, all of the personality items that reflected masculine or feminine stereotypes showed the change of endorsement when the gender-associated stereotype was introduced.

The anticipated relationship of competent and warm stereotype-coded items and the expected associated gender did not hold. Contrary to the association in the stereotype literature, women were more likely to endorse the competent item, whereas men were more likely to endorse the not competent item. This finding may lend that competence is more of a neutral characteristic, such that men and women associate themselves with being competent. Competence has traditionally been assigned to men, however, the association of women and competence has increased in current consensual beliefs (Eagly et al., 2020).

This section of this research looked to dive deeper into how language and changes in words can impact the item endorsement by gender. The findings here illustrated that stereotypes can be present in personality items and in some instances, it can impact whether men or women endorse the item. The effect of stereotypes in personality items provides insight to the implications and future directions of this research.

**Theoretical Implications**

The goal of the research was to identify if personality items include content that is construct-irrelevant to the latent personality trait, resulting gender differences in endorsement. As there was a small magnitude shown here, there is still some evidence that we should look further into how we design and evaluate personality items. In this research, it was not the expectation that all personality items would show DIF or that all personality assessments are overwhelmed with stereotypes. Although many factors may contribute to gender differences in personality. As stereotypes are present in society, it is possible that these gender-based expectations are real differences due to socialization. However, the pattern of results presents an opportunity to assess how we measure personality and construct personality items. Here, there may be an implication to a fundamental limitation of how personality is defined and represented in personality assessments.

Personality items often use language rooted in adjectives and there are even personality assessments that use single-item adjective rating scales (Goldberg, 1990; Paunonen & Hong, 2015). Personality traits are defined through the lexical hypothesis and are a collection of phrases and sentences (Saucier & Goldberg, 1996). These phrases and sentences are often adjectives as adjectives allow a representative sample of personality attributes, desirable and undesirable (Saucier & Goldberg, 1996). As personality items are deeply rooted in adjectives, they embed all of what language reflects, including social stereotypes. Thus, stereotypes can possibly influence the measurement of personality.

When defining personality, the fundamental goal is to accurately describe actual or potential behavior (Saucier & Srivastava, 2015). Recognizing that behavior is the critical focus of personality, a potential direction when constructing personality items would be to develop them to reflect traits through behavior, more so than traits through adjectives. With the

theoretical and measurement implications, we should consider how we measure personality in a way that is less adjective based or at least in a way that more gender-neutral adjectives are used. Although this paper focused on gender stereotypes, this can also apply to cultural, racial, and other stereotypes as well.

**Practical Implications**

The culmination of these findings allows for practical insights into personality items, assessments, and application. This research showed that IPIP items are reliable across different samples and that if the item uses gendered language, stereotypes can be present and impact the endorsement. Personality items can be influenced by the words in the item as to be expected. Items are written to reflect a specific personality trait and subtrait. While the language in the item is purposeful, this research shows that personality items can reflect gender stereotypes that can thus impact endorsement of the items. The presence of stereotype content can influence the way in which men and women endorse the item in alignment to the associated stereotype, even as their placement on the latent personality trait would indicate no differences. As the IPIP did not show many items exhibiting invariance, exercising a deeper level of cautiousness in making comparisons of means between gender is important.

The results of this research also provides insight to the possible preparation and approach needed when developing personality assessments. For many types of assessment, sensitivity panels and diversity panels have been put into place to make sure that we are not including content or language that may be perceived or responded to differentially across groups. Using cognitive ability assessments as an example, these assessments have sometimes been found to have measurement impact (e.g. construct bias, method bias, item bias) on certain groups (Reynolds & Suzuki, 2013). To address these concerns, there have been processes established to

minimize this impact, such as a tryout sample or norming sample. A tryout sample is when the test author or publisher administers the items to a group of people (Reynolds & Suzuki, 2013). Following this exercise, the remaining items are administered to a large, diverse group of people. The norming sample reflects the important characteristics of the population that the test is expected to be administered to. It can often include racial and ethnic group representation to understand the impact of group membership on the test being developed. Related to personality assessments, when constructing personality items, an opportunity to continue to improve personality items is to use these sampling procedures to pre-assess items for stereotypes. Early identification in the development stage can positively impact the usefulness and applicability of personality assessment. Thus, it is important to include these types of processes when designing or collecting personality or non-cognitive items.

Additionally, from the psychometric lens, this research also highlights an opportunity for organizations to be informed consumers when identifying and selecting personality assessments. As organizations work with vendors to employ personality assessments for different organizational needs, such as selection, coaching, job analysis, it is important to ask if the personality assessment or tool has been reviewed for any potential construct-irrelevant variance and if one can be confident that individual differences that will be found are valid or real. Engaging in these questions and review of personality assessment can help ensure that the tools being used for critical talent decisions are valid, reliable, and fair.

This research is also impactful in the leadership development and assessment space. Personality inventories are often used to understand leaders and/or managers' competencies, leadership traits, working styles, etc. When dissecting and interpreting those results, sometimes single items on the assessment are reviewed to identify a specific behavior. Seeing that

individual items can show DIF even if the full scale does not, there should be some hesitation to interpreting single items. The impact of stereotypes on items further supports the analysis of these measures at the scale level. At the aggregate level, with the number of items showing DIF, the integrity of the scale or items can still be intact.

**Limitations and Future Directions**

One important limitation of this research is the limited sample size in the IRT analysis. IRT analyses often require a large sample size (around 1,000) to accurately estimate parameters of ability (Hambleton, Swaminathan, & Rogers, 1991). Varying research states this number can vary from 250 to 500 based on the IRT model being used (Goldman & Raju, 1986; Guyer & Thompson, 2011; Thissen & Wainer, 1982). Even with the limitation of sample size, the fit of the model was acceptable and assumptions for the analysis were adequate to complete the analyses. Moreover, in combination with the Mantel procedure, this analysis allows for convergence when determining the presence of DIF. A replication of this study with a larger sample size is a necessary next step. Additionally, future research should also focus on diverse samples to investigate the impact of personality items on different ages, races, cultures, etc.

Another limitation of this study is that the complete measure of personality traits was not included. Without the complete personality trait, there was an impact to being able to look in detail at the interaction between stereotypes and personality characteristics. The purpose of this research was to show the impact of stereotypes on responses to personality items. As the original items were rewritten, especially the opposite stereotype version, it is possible that the new item was adjusted from the personality construct it originally reflected. Moving forward in this line of research, future studies can dive into measuring personality traits and if items that reflect stereotypes are measuring the intended personality trait. Additionally, with testing different

versions of the item, further research should dive into construct validity and assess if the different version continues to reflect the personality trait as measured.

One other potential limitation to this study is related to the coding identification of stereotypes. In the first study, the coding procedure used terms that describe the classification of gender-based stereotypes. While these terms were fully explained and illustrated to coders, the use of different, common stereotypical traits may have yielded in a greater percentage of items identified as reflecting common gender stereotypes. As this alternative approach may have resulted in increased identification, the categories used in the stereotype literature provided common terminology and a clearer theoretical linkage.

Looking to the future direction of this research, some next steps are to examine different personality assessments and different response options. This research looked at one set of personality items which is a limitation to the current study. However, the IPIP is a very common personality assessment but there are various personality assessments comprised of different types of items. To continue the research and the understanding of the presence of stereotypes, it is of interest to review popular personality assessments, such as Hogan, DISC, CPI, Myers-Briggs, etc.

Similar to the personality measures, there is also an opportunity to examine the ways in which participants are asked to respond to items. This study looked at self-report items on a single measure. However, there are various assessment methods of personality, like projective tests and objective measures. Self-report assessments also vary from forced choice statements, adjective checklists, or open-ended descriptions of self (Paunonen & Hong, 2015). The way in which participants are asked to select or respond to personality items could increase the prevalence or impact that a stereotype embedded in the item may have. Further research should

look into if different response options or response modalities influence the impact of stereotypes in personality items.

An additional exploration of this research is connected to the faking literature in the personality domain. Many studies have looked at how individuals can attempt to fake their responses and the impact that is has on the results of the assessments. Item response theory is often the methodology to identify the presence of faking. Related to this research, a potential factor to explain gender-based differences on personality could be faking such that individuals are deliberately endorsing certain personality items to match the expectation of their gender. Alternatively, it is possible that participants are experiencing stereotype threat, such that, participants are "faking" responses to endorse a trait that is opposite of their associated gender stereotype. Looking into this connection will provide greater insight into the presence and influence of stereotypes.

Future directions should also lend to the impact of context or other potential moderators. The person-situation interaction is a prominent component in the personality domain. This interaction focuses on the strength of the environment and how strongly people are influenced by it. These theories look to identify the need to incorporate the influence of person characteristics and situation characteristics in the prediction of responses (Moskowitz & Fournier, 2014). Related to this specific research, we should continue to explore if personality assessments taken in specific contexts, at work versus at home, exacerbate or minimize the impact of stereotypes on personality items. Research continues to show that traditional gender roles exist in the home even with expansion of roles in the workforce (Bianchi, Sayer, Milkie, & Robinson, 2012; Pew Research Survey, 2012; Rao, 2019). The presence and impact of stereotypes in personality items may be contingent on the domain in which the assessment is given. This example provides

insight into how additional research looking into moderators can expound our understanding of the relationship between personality items and stereotypes.

**Conclusion**

An ideal personality assessment is one in which the probability of endorsing an item depends only on the individual's level on the personality trait, factoring in valid measurement components. When there is an indication of measurement invariance or DIF, this would suggest there are variables outside of the construct impacting the endorsement, such as gender-based stereotypes, as examined here. If DIF is present, the measurement needs to be reassessed to assure that it is fairly measuring across groups. This research adds measurement as an important explanation of gender differences on personality items. The current study does not necessarily provide the answer to how to remove this impact nor suggests that there are not real differences in personality by gender, but it does elucidate that we should be more thoughtful on how we have been constructing personality items and the adjectives that are used to reflect personality traits may be perceived and influence the direction of endorsement by different groups. Exploring all of the potential explanations and critically reviewing personality inventories for construct irrelevant variance will assure that we are using measures that are valid, reliable, and fair. As the literature on personality and personality assessments continues to grow, these studies contribute to the opportunity to further understand how gender differences present themselves and the potential impact to the organizational applications for which personality assessments are used.

## Appendix A

## The International Personality Item Pool

On the following pages, there are phrases describing people's behaviors. Please use the rating scale next to each phrase to describe how accurately each statement describes **you**. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, and then fill in the bubble that corresponds to the accuracy of the statement.

Response Options:

1- Very Inaccurate 2-Moderately Inaccurate 3-Neither Inaccurate nor Accurate

4- Moderately Accurate 5- Very Accurate

1. Excel in what I do.
2. Become overwhelmed by events.
3. Can handle a lot of information.
4. Stick to the rules.
5. Sympathize with the homeless.
6. Believe in one true religion.
7. Leave my belongings around.
8. Make people feel welcome.
9. Distrust people.
10. Cheer people up.
11. Love surprise parties.
12. Jump into things without thinking.
13. Believe that others have good intentions.
14. Cheat to get ahead.
15. Have a good word for everyone.

16. Believe that we should be tough on crime.
17. Am not easily affected by my emotions.
18. Often feel uncomfortable around others.
19. Love flowers.
20. Am indifferent to the feelings of others.
21. Can manage many things at the same time.
22. Am not highly motivated to succeed.
23. Turn my back on others.
24. Try to lead others.
25. Have a high opinion of myself.
26. Try not to think about the needy.
27. Am easy to satisfy.
28. Like to visit new places.
29. Experience very few emotional highs and lows.
30. Can handle complex problems.
31. Believe that people are basically moral.
32. Do things I later regret.
33. Want everything to be “just right.”
34. Handle tasks smoothly.
35. Am able to control my cravings.
36. Feel that my life lacks direction.
37. Use others for my own ends.
38. Love action.
39. Don’t like crowded events.
40. Panic easily.
41. Love life.
42. Trust others
43. Am easily intimidated.
44. Don’t know why I do some of the things I do.
45. Demand quality.
46. Stick to my chosen path.

47. Pay my bills on time.
48. Rarely overindulge.
49. Am very pleased with myself.
50. Carry out my plans.
51. Interested in many things.
52. Enjoy being part of a group.
53. Have a low opinion of myself.
54. See beauty in things that others might not notice.
55. Willing to try anything once.
56. Like order.
57. Break my promises.
58. Rarely notice my emotional reactions.
59. Avoid mistakes.
60. Seldom get mad.
61. Remain calm under pressure.
62. Am relaxed most of the time.
63. Am not really interested in others.
64. Often eat too much.
65. Prefer variety to routine.
66. Avoid philosophical discussions.
67. Keep others at a distance.
68. Believe that I am better than others.
69. Dislike changes.
70. Take no time for others.
71. Love to help others.
72. Often feel blue.
73. Can’t stand weak people.
74. Don’t like to draw attention to myself.
75. Do not like concerts.
76. Do a lot in my spare time.
77. Have difficulty understanding abstract ideas.

78. Am a creature of habit.
79. Feel desperate.
80. Am not bothered by messy people.
81. Look at the bright side of life.
82. Get overwhelmed by emotions.
83. Am not interested in other people's problems.
84. Use flattery to get ahead.
85. Spend time reflecting on things.
86. Believe in the importance of art.
87. Believe laws should be strictly enforced.
88. Do just enough work to get by.
89. Indulge in my fantasies.
90. Like to begin new things.
91. Get caught up in my problems.
92. Wait for others to lead the way.
93. Anticipate the needs of others.
94. Have a vivid imagination.
95. Have a lot of fun.
96. Turn plans into actions.
97. Find it difficult to get down to work.
98. Tell the truth.
99. Feel comfortable with myself.
100. Get angry easily.
101. Get chores done right away.
102. Waste my time.
103. Enjoy the beauty of nature.
104. Feel others' emotions.
105. Like a leisurely lifestyle.
106. Do things according to a plan.
107. Seldom feel blue.
108. Rush into things.

109. Can't stand confrontations.
110. Radiate joy.
111. Seek to influence others.
112. Have little to say.
113. Am passionate about causes.
114. Am wary of others.
115. Get others to do my duties.
116. Like to act on a whim.
117. Know how to cope.
118. Believe in human goodness.
119. Dislike myself.
120. Try to follow the rules.
121. Act without thinking.
122. Find it difficult to approach others.
123. Am not bothered by difficult social situations.
124. React quickly.
125. Experience my emotions intensely.
126. Obstruct others' plans.
127. Pretend to be concerned for others.
128. Listen to my conscience.
129. Get back at others.
130. Would never cheat on my taxes.
131. Dislike talking about myself.
132. Trust what people say.
133. Easily resist temptations.
134. Am not easily disturbed by events.
135. Am calm even in tense situations.
136. Am hard to get to know.
137. Rarely get irritated.
138. Don't worry about things that have already happened.
139. Am always prepared.

140. Enjoy wild flights of fantasy.
141. Have a rich vocabulary.
142. Enjoy examining myself and my life.
143. Contradict others.
144. Do not enjoy going to art museums.
145. Like to solve complex problems.
146. Seldom joke around.
147. Love a good fight.
148. Avoid crowds
149. Seldom toot my own horn.
150. Dislike loud music.
151. Set high standards for myself and others.
152. Let things proceed at their own pace.
153. Am afraid that I will do the wrong thing.
154. Insult people.
155. Can talk others into doing things.
156. Fear for the worst.
157. Come up with good solutions.
158. Choose my words with care.
159. Complete tasks successfully.
160. Have difficulty imagining things.
161. Adapt easily to new situations.
162. Avoid contacts with others.
163. Often forget to put things back in their proper place.
164. Am always on the go.
165. Believe that people are essentially evil.
166. Can't make up my mind.
167. Love order and regularity.
168. Believe people should fend for themselves.
169. Am not easily bothered by things.
170. Do not like poetry.

171. Love large parties.
172. Enjoy being part of a loud crowd.
173. Stumble over my words.
174. Don't like the idea of change.
175. Believe that we coddle criminals too much.
176. Have little to contribute.
177. Don't understand things.
178. Laugh my way through life.
179. Consider myself an average person.
180. Seldom get emotional.
181. Act comfortably with others.
182. Keep my cool.
183. Break rules.
184. Am often down in the dumps.
185. Am always busy.
186. Misrepresent the facts.
187. Express childlike joy.
188. Get stressed out easily.
189. Take advantage of others.
190. Hold back my opinions.
191. Know how to get around the rules.
192. Get irritated easily.
193. Would never go hang gliding or bungee jumping.
194. Avoid difficult reading material.
195. Prefer to stick with things that I know.
196. Make friends easily.
197. Like to take my time.
198. Am not easily amused.
199. Am concerned about others.
200. Laugh aloud.
201. Do more than what's expected of me.

202. Tend to vote for liberal political candidates.
203. Am afraid to draw attention to myself.
204. Make rash decisions.
205. Warm up quickly to others.
206. Rarely complain.
207. Don't see the consequences of things.
208. Like to stand during the national anthem.
209. Feel comfortable around people.
210. Know how to get things done.
211. Love to daydream.
212. Amuse my friends.
213. Postpone decisions.
214. Put little time and effort into my work.
215. Seldom daydream.
216. Lose my temper.
217. Never spend more than I can afford.
218. Have frequent mood swings.
219. Leave a mess in my room.
220. Love excitement.
221. Believe that criminals should receive help rather than punishment.
222. Go on binges.
223. Try to understand myself.
224. Love to read challenging material.
225. Dislike being the center of attention.
226. Love to eat.
227. Yell at people.
228. Do not like art.
229. Prefer to be alone.
230. Enjoy being reckless.
231. Go straight for the goal.
232. Tend to dislike soft-hearted people.

233. Never splurge.
234. Have difficulty starting tasks.
235. Am not interested in theoretical discussions.
236. Know the answers to many questions.
237. Seek quiet.
238. Hold a grudge.
239. Do not have a good imagination.
240. Make myself the center of attention.
241. Get upset easily.
242. Hate to seem pushy.
243. Only feel comfortable with friends.
244. Like to take it easy.
245. Am not bothered by disorder.
246. Talk to a lot of different people at parties.
247. Am not embarrassed easily.
248. Like to tidy up.
249. Boast about my virtues.
250. Believe in an eye for an eye.
251. Suspect hidden motives in others.
252. Am not interested in abstract ideas.
253. Work hard.
254. Am comfortable in unfamiliar situations.
255. Readily overcome setbacks.
256. Get to work at once.
257. Seek danger.
258. Tend to vote for conservative political candidates.
259. Think that all will be well.
260. Believe that there is no absolute right or wrong.
261. Have a sharp tongue.
262. Think highly of myself.
263. Am afraid of many things.

264. Believe that too much tax money goes to support artists.
265. Am often in a bad mood.
266. Want to be left alone.
267. Keep my promises.
268. Don't understand people who get emotional.
269. Make people feel uncomfortable.
270. Seek adventure.
271. Do the opposite of what is asked.
272. React slowly.
273. Keep in the background.
274. Plunge into tasks with all my heart.
275. Am attached to conventional ways.
276. Suffer from others' sorrows.
277. Dislike new foods.
278. Feel that I'm unable to deal with things.
279. Value cooperation over competition.
280. Look down on others.
281. Am not easily annoyed.
282. Seldom get lost in thought.
283. Enjoy thinking about things.
284. Put people under pressure.
285. Like music.
286. Misjudge situations.
287. Worry about things.
288. Involve others in what I am doing.
289. Take charge.
290. Often make last-minute plans.
291. Like to get lost in thought.
292. Am able to stand up for myself.
293. Do not enjoy watching dance performances.
294. Do crazy things.

295. Need a push to get started.
296. Am sure of my ground.
297. Feel sympathy for those who are worse off than myself.
298. Act wild and crazy.
299. Start tasks right away.
300. Take control of things.

Appendix B

Coding Directions

**Instructions:**

The purpose of this coding exercise is to categorize a select set of personality items. You will be reading each item and assessing which category the item best fits. In order to understand the selections, these definitions for these anchors are:

***Agentic/ Communal Scale***

Agentic: Agentic classifications are associated with assertion and control. Traits associated with agentic beliefs are aggressive, ambitious, dominant, self-confident, forceful, self-reliant, self-sufficient, and individualistic.

Communal: Communal is the dimension that is characterized as compassionate treatment; these traits are commonly affectionate, helpful, friendly, kind, sympathetic, interpersonally sensitive, gentle, and soft-spoken.

Using the definitions and the coding training guide, read each item and rate it on each 1 to 5 scale with 1 representing not at all agentic/communal to 5 representing very agentic/communal.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Not at all agentic | | Neutral | | Very agentic |

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Not at all communal | | Neutral | | Very communal |

***Warmth/Competent Scale***

Competence: Competence has been synonymous with capability, skill, and talent. Traits associated with competence are clever, competent, creative, dominance, efficient, foresighted, ingenious, intelligent, knowledgeable.

Warmth: Warmth includes emotionality and empathy. The warmth component is associated with traits such as morality, trustworthiness, sincerity, kindness, and friendliness.

When coding the item, identify if the item you are reading reflects the presence of a competent quality or warm quality.

Using the definitions and the coding training guide, read each item and rate it on each 1 to 5 scale with 1 representing not at all warm or not at all competent to 5 representing mostly warm or mostly competent.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Not at all Warm | | Neutral | | Very Warm |

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Not at all Competent | | Neutral | | Very Competent |

***Masculine/Feminine Scale***

Masculine: Masculine is defined as a set of attributes, behaviors, and roles that are traditionally associated with men, which includes strength, courage, and aggression. Traits associated with masculinity are courage, risk-taker, dominance, ambitious, independence, assertive, competitive, rational, and emotional control.

Feminine: Feminine is defined as a set of attributes, behaviors, and roles that are traditionally associated with women, which includes gentleness, empathy, and sensitivity. Traits associated with femininity are gentleness, modestly, humility, sacrifice, supportiveness, caring, compassion, affectionate, nurturance, and sensitivity.

When coding the item, identify if the item you are reading reflects the presence of a competent quality or warm quality.

Using the definitions and the coding training guide, read each item and rate it on each 1 to 5 scale with 1 representing masculine to 5 representing feminine.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Masculine | | Neutral | | Feminine |

Appendix C:
Coding Training Guide

This training guide is created to provide a framework of which personality items are fit in the appropriate categories. The first set of items selected will represent examples of clearly agentic/communal/warmth/competent items. The second set of items here will represent examples of items that are more ambiguous but still fit in the category.

Below are examples of personality items that are very clearly and ambiguously agentic.

**Cleary agentic:**
Act as a leader.
Automatically take charge.
Act impulsively when something is bothering me.
Can control the outcome of events.

**Ambiguous- agentic:**
Accomplish my work on time.
Can handle a lot of information.
Deal efficiently with practical matters.
Do not feel close to others.

Below are examples of personality items that are clearly communal and ambiguously communal.

**Clearly communal:**
Accept apologies easily.
Approach others in a positive manner.
Go out of my way for others.
Speak softly.

**Ambiguous- communal:**
Acknowledge others' accomplishments
Can be relied upon by others.
Believe the poor deserve our sympathy.
Give everyone a chance.

Below are examples of personality items that are clearly competent and ambiguously competent.

**Clearly competent:**
Handle tasks smoothly.
Get things done quickly.
Know how to apply my knowledge.
Know my strengths

**Ambiguous-competent:**

Have a very good imagination.
Have read a lot.
Investigate all capabilities.
Like to read.

Below are examples of personality items that are clearly warm and ambiguously warm.

**Very warm:**
Make people feel welcome.
Make people feel at ease.
Reassure others.
Radiate joy.

**Ambiguous- warm**:
Never resent being asked to do a favor for someone.
Rarely or never raise my voice in anger.
Would find it distressing to someone sleeping on the streets.
Am a law-abiding citizen.

Appendix D
Pilot Study Survey

Masculine/Feminine

The purpose of this study is to read personality items and assess the item for the presence of qualities of masculinity and femininity. You will be reading each item and assessing if the item contains content or behavior that is connected to masculinity or femininity.

As you read the item, assess if the item or behavior is more masculine, feminine, or neutral. In order to understand the options, these definitions for these categories are:

Masculine: Masculine is defined as a set of attributes, behaviors, and roles that are traditionally associated with men, which includes strength, courage, and aggression. Traits associated with masculinity are:

- courage,
- risk-taker,
- dominance,
- ambitious,
- independence,
- assertive,
- competitive,
- rational,
- emotional control

Feminine: Feminine is defined as a set of attributes, behaviors, and roles that are traditionally associated with women, which includes gentleness, empathy, and sensitivity. Traits associated with femininity are:

- gentleness,
- modestly,
- humility,
- sacrifice,
- supportiveness,
- caring,
- compassion,
- affectionate,
- nurturance,
- sensitivity

Using the definitions read each item and rate it on a scale of masculine and feminine. If the item does not reflect either trait, select the neutral option.

Response options: 1-Masculine, 2- Neutral, 3-Feminine

1a. Love flowers.
1b. Love trees.
1c. Love woodcarving.
2a. Get overwhelmed by emotions.
2b. Get impacted by moods.
2c. Not overwhelmed by emotions.
3a. Experience my emotions intensely.
3b. Experience my emotions.
3c. Don’t notice my emotions.
4a. Am not easily affected by my emotions.
4b. Am not easily impacted by my affect.
4c. Am easily affected by my emotions.
5a. Love a good fight.
5b. Don’t mind confrontation.
5c. Dislike fights.
6a. Break rules.
6b. Defy expected norms.
6c. Follow rules.
7a. Get upset easily.
7b. Can get upset.
7c. Don’t like to get upset.
8a. Don’t understand people who get emotional.
8b. Don’t understand people who are affected by their feelings.
8c. Sympathize with those who get emotional.

Agentic/Communal

The purpose of this study is to read personality items and assess the item for the presence of qualities of agentic and communal. You will be reading each item and assessing if the item contains content or behavior that is connected to agentic or communal content.

As you read the item, assess if the item or behavior is more agentic, communal, or neutral. In order to understand the options, these definitions for these categories are:

Agentic: Agentic classifications is defined as assertion and control. Traits associated with agentic are:

- aggressive,
- ambitious,
- dominant,
- self-confident,
- forceful,
- self-reliant,
- self-sufficient
- individualistic.

Communal: Communal is defined as the compassionate treatment of others. Traits associated with communal are:

- affectionate,
- helpful,
- friendly,
- kind,
- sympathetic,
- interpersonally sensitive,
- gentle,
- soft-spoken.

Using the definitions read each item and rate it on a scale of agentic and communal. If the item does not reflect either trait, select the neutral option.

Response Options: 1-Agentic, 2-Neutral, 3-Communal

1a. Hate to seem pushy.
1b. Hate to seem forward.
1c. Comfortable being pushy.
2a. Value cooperation over competition.
2b. Value cooperation and competition.
2c. Value competition over cooperation.
3a. Have a high opinion of myself.
3b. Have high self-esteem.
3c. Have a modest opinion of myself.
4a. Remain calm under pressure.
4b. Remain at ease under pressure.

4c. Help others stay calm under pressure.
5a. Get chores done right away.
5b. Get chores done immediately.
5c. Get mine and others chores done right away.
6a. Try to understand myself.
6b. Try to gain insight into myself.
6c. Interested in gaining insight into others.
7a. Think highly of myself.
7b. Think confidently of myself.
7c. Think highly how I care for others.

Warm

The purpose of this study is to read personality items and assess the item for the presence of qualities of warmth. You will be reading each item and assessing if the item contains content or behavior that is connected to warm content.

Warmth: Warmth is defined as emotionality and empathy. Traits associated with warmth are

- morality,
- trustworthiness,
- sincerity,
- kindness,
- friendliness

Using the definition, assess if the item or behavior is not warm, neutral, or warm.
Response options 1- Not Warm, 2- Neutral, 3- Warm

1a. Seldom get mad.
1b. Seldom get upset.
1c. Get mad often.
2a. Experience my emotions intensely.
2b. Experience my emotions.
2c. Don't experience emotions.
3a. Am relaxed most of the time.
3b. Am sometimes relaxed.
3c. Am often anxious.
4a. Rarely complain.
4b. Sometimes I complain.
4c. Often complain.
5a. Try to understand myself.
5b. Sometimes try to understand myself.
5c. Not interested in understanding myself.

Competent

The purpose of this study is to read personality items and assess the item for the presence of qualities of competence. You will be reading each item and assessing if the item contains content or behavior that is connected to competence.

Competence: Competence is defined as capability, skill, and talent. Traits associated with competence are:

- clever,
- competent,
- creative,
- dominance,
- efficient,
- foresighted,
- ingenious,
- intelligent,
- knowledgeable

Using the definition, assess if the item or behavior is not competent, neutral, or competent. Response options 1- Not competent, 2-Neutral, 3-Competent

1a. Like order.
1b. Don't mind order.
1c. Prefer being disorganized.
2a. Easily resist temptations.
2b. Somewhat easily resist temptations.
2c. Unable to resist temptations.
3a. Am calm in even tense situations.
3b. Somewhat calm in tense situations.
3c. Am unable to adapt in tense situations.
4a. Adapt easily to new situations.
4b. Adjust to new situations.
4c. Can't successfully adapt to new situations.
5a. Know the answers to many questions.
5b. Am knowledgeable about some things.
5c. Don't often know the answers.

Appendix E
Manipulated Personality Items

Please complete this personality inventory as honestly as you can. The results will be completely anonymous and used for research purposes only.

Below, you will see phrases describing people's behaviors. Please use the rating scale next to each phrase to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future.

It is very important that you respond to this survey by describing yourself as you really are and not as you want to be or as you want others to see you.

Select the response that describes yourself as you honestly see yourself. Your responses will be kept in absolute confidence so that you may respond honestly.

Read each statement carefully, and then fill select the option that best describes you.

Response Options:

1- Very Inaccurate 2-Moderately Inaccurate 3-Neither Inaccurate nor Accurate

4- Moderately Accurate 5- Very Accurate

1. Get overwhelmed by emotions.
2. Get impacted by moods.
3. Not overwhelmed by emotions.
4. Experience my emotions intensely.
5. Experience my emotions.
6. Don't notice my emotions.
7. Am not easily affected by my emotions.
8. Sometimes act based on my emotions.
9. Am easily affected by my emotions.
10. Don't understand people who get emotional.
11. Don't understand people who are affected by their feelings.
12. Sympathize with those who get emotional.
13. Value cooperation over competition.
14. Value cooperation and competition.
15. Value competition and cooperation.
16. Remain calm under pressure.
17. Remain at ease under pressure.
18. Help others stay calm under pressure.
19. Get chores done right away.
20. Get chores done.
21. Get mine and others chores done right away.

22. Think highly of myself.
23. Think well of myself.
24. Think highly of how I care for others.
25. Am relaxed most of the time.
26. Am sometimes relaxed.
27. Am often anxious.
28. Rarely complain.
29. Sometimes I complain.
30. Often complain.
31. Like order.
32. Don't mind order.
33. Dislike order.
34. Easily resist temptations.
35. Somewhat easily resist temptations.
36. Unable to resist temptations.
37. Am calm in even tense situations.
38. Somewhat calm in even tense situations.
39. Am not calm in tense situations.
40. Adapt easily to new situations.
41. Adapt to new situations.
42. Can't successfully adapt to new situations.
43. Know the answers to many questions.
44. Am able to answer some questions.
45. Often don't know the answers to many questions.
46. Excel in what I do.
47. Excel in some things that I do.
48. Don't excel at anything I do.
49. Radiate joy.
50. Occasionally joyous.
51. Don't often feel joy.
52. Trust what people say.
53. Sometimes trust what people say.
54. Never trust what people say.
55. Do more than what's expected of me.
56. Do what is generally expected of me.
57. Do less than what is expected of me.
58. Amuse my friends.
59. Entertain others.
60. Don't amuse my friends.
61. Think that all will be well.
62. Often think that all will end up ok.
63. Hardly ever think that all will be well.
64. Involve others in what I am doing.

65. Sometimes involve others in what I am doing.
66. Prefer to do things alone.

Table 1

*Sub-factors for each Dimension of the Five Factor Model of Personality*

| Neuroticism | Extraversion | Openness to Experience | Agreeableness | Conscientiousness |
|---|---|---|---|---|
| Anxiety | Warmth | Fantasy | Trust | Competence |
| Hostility | Gregariousness | Aesthetics | Straightforwardness | Order |
| Depression | Assertiveness | Feelings | Altruism | Dutifulness |
| Self-consciousness | Activity | Actions | Compliance | Achievement Striving |
| Impulsiveness | Excitement-Seeking | Ideas | Modesty | Self-Discipline |
| Vulnerability | Positive Emotions | Values | Tender-mindedness | Deliberation |

Table 2

*Mean z-score Differences (d) Between Women and Men on Revised NEO Personality Inventory (NEO-PI-R) Facets in the United States*

| NEO-PI-R facet | U.S. Adults |
|---|---|
| N1: Anxiety | .40*** |
| N2: Angry Hostility | .09 |
| N3: Depression | .24*** |
| N4: Self-Consciousness | .30*** |
| N5: Impulsiveness | .23*** |
| N6: Vulnerability | .44*** |
| E1: Warmth | .33*** |
| E2: Gregariousness | .21*** |
| E3: Assertiveness | -.19** |
| E4: Activity | .11* |
| E5: Excitement Seeking | -.31*** |
| E6: Positive Emotions | .29*** |
| O1: Fantasy | -.16** |
| O2: Aesthetics | .34*** |
| O3: Feelings | .28** |
| O4: Actions | .19*** |
| O5: Ideas | -.32*** |
| O6: Values | -.07 |
| A1: Trust | .19** |
| A2: Straightforwardness | .43*** |
| A3: Altruism | .43*** |
| A4: Compliance | .38*** |
| A5: Modesty | .38*** |
| A6: Tender-Mindedness | .31*** |
| C1: Competence | -.20*** |

| | |
|---|---|
| C2: Order | .05 |
| C3: Dutifulness | .00 |
| C4: Achievement Striving | .08 |
| C5: Self-Discipline | -.02 |
| C6: Deliberation | -.12 |

*$p$ < .05, ** $p$ <.01, *** $p$<.001

Table 3

*Gender-Related Trait Items*

| Communion | Agency |
|---|---|
| Emotional | Independent |
| Able to devote self to others | Active |
| Gentle | Competitive |
| Helpful to others | Can make decisions easily |
| Kind | Never gives up |
| Aware of others' feelings | Self-confident |
| Understanding of others | Feels superior |
| Warm in relations to others | Stands up well under pressure |

Table 4

*Means and Variances of IPIP Items*

| Item | Factor | M | $S^2$ |
|---|---|---|---|
| 1 | C | 3.88 | .79 |
| 7 | C | 3.43 | 1.69 |
| 12 | C | 3.22 | 1.48 |
| 22 | C | 4.13 | 1.14 |
| 33 | C | 3.94 | 1.02 |
| 34 | C | 3.61 | .72 |
| 45 | C | 4.01 | .80 |
| 46 | C | 3.49 | 1.03 |
| 47 | C | 4.10 | 1.13 |
| 50 | C | 3.77 | .86 |
| 56 | C | 3.72 | 1.19 |
| 57 | C | 3.89 | .97 |
| 59 | C | 3.48 | .99 |
| 80 | C | 3.49 | 1.32 |
| 88 | C | 3.27 | 1.41 |
| 96 | C | 3.68 | .933 |
| 97 | C | 3.21 | 1.33 |
| 98 | C | 3.93 | .85 |
| 101 | C | 2.62 | 1.28 |
| 102 | C | 3.06 | 1.38 |
| 106 | C | 3.56 | .88 |
| 108 | C | 3.01 | 1.26 |
| 115 | C | 3.77 | 1.16 |
| 116 | C | 3.27 | .93 |
| 120 | C | 3.98 | .82 |
| 121 | C | 3.44 | 1.34 |
| 128 | C | 3.78 | .68 |
| 139 | C | 3.32 | 1.03 |
| 151 | C | 3.97 | 1.00 |
| 157 | C | 3.89 | .63 |
| 158 | C | 3.45 | 1.00 |
| 159 | C | 3.82 | .69 |
| 163 | C | 3.36 | 1.61 |
| 167 | C | 3.77 | 1.21 |
| 176 | C | 3.79 | .97 |
| 177 | C | 3.83 | .94 |
| 183 | C | 3.58 | 1.15 |
| 186 | C | 3.62 | .90 |
| 201 | C | 3.58 | 1.02 |

| | | | |
|---|---|---|---|
| 204 | C | 3.21 | 1.08 |
| 207 | C | 3.73 | 1.04 |
| 210 | C | 3.97 | .59 |
| 213 | C | 3.06 | 1.23 |
| 214 | C | 3.83 | 1.07 |
| 219 | C | 3.01 | 1.79 |
| 231 | C | 3.82 | .88 |
| 234 | C | 3.10 | 1.33 |
| 245 | C | 3.26 | 1.18 |
| 248 | C | 3.41 | 1.03 |
| 253 | C | 4.01 | .89 |
| 256 | C | 3.16 | 1.15 |
| 267 | C | 3.94 | .908 |
| 271 | C | 3.97 | .87 |
| 274 | C | 3.37 | .942 |
| 286 | C | 3.23 | 1.03 |
| 290 | C | 2.54 | 1.30 |
| 294 | C | 3.12 | 1.47 |
| 295 | C | 3.13 | 1.32 |
| 296 | C | 3.63 | .80 |
| 299 | C | 3.05 | .80 |
| | | | |
| 10 | E | 3.92 | .83 |
| 11 | E | 3.62 | 1.57 |
| 18 | E | 3.24 | 1.29 |
| 21 | E | 3.73 | .89 |
| 24 | E | 3.47 | 1.07 |
| 38 | E | 3.87 | .88 |
| 39 | E | 3.16 | 1.65 |
| 41 | E | 3.96 | 1.19 |
| 52 | E | 3.78 | 1.19 |
| 55 | E | 3.62 | 1.41 |
| 63 | E | 3.75 | 1.08 |
| 67 | E | 3.48 | 1.21 |
| 74 | E | 2.87 | 1.67 |
| 76 | E | 3.09 | 1.35 |
| 81 | E | 3.67 | 1.23 |
| 92 | E | 3.40 | 1.27 |
| 95 | E | 3.75 | 1.18 |
| 105 | E | 2.11 | .99 |
| 110 | E | 3.36 | .97 |
| 111 | E | 3.57 | 1.09 |
| 112 | E | 3.44 | 1.46 |
| 124 | E | 3.47 | .85 |
| 136 | E | 3.26 | 1.54 |
| 146 | E | 3.62 | 1.55 |

| | | | |
|---|---|---|---|
| 148 | E | 3.20 | 1.48 |
| 150 | E | 3.88 | 1.39 |
| 152 | E | 2.52 | .93 |
| 155 | E | 3.48 | .98 |
| 162 | E | 3.77 | 1.12 |
| 164 | E | 3.31 | 1.08 |
| 171 | E | 3.54 | 1.63 |
| 172 | E | 3.19 | 1.68 |
| 178 | E | 3.02 | 1.34 |
| 181 | E | 3.70 | .84 |
| 185 | E | 3.34 | 1.33 |
| 187 | E | 3.35 | 1.21 |
| 190 | E | 3.25 | 1.39 |
| 193 | E | 3.41 | 2.31 |
| 196 | E | 3.49 | 1.07 |
| 197 | E | 2.41 | .96 |
| 198 | E | 3.57 | 1.18 |
| 200 | E | 3.73 | 1.40 |
| 205 | E | 3.44 | .78 |
| 209 | E | 3.57 | 1.03 |
| 212 | E | 3.84 | .84 |
| 220 | E | 4.14 | .709 |
| 229 | E | 3.23 | 1.40 |
| 230 | E | 2.36 | 1.14 |
| 237 | E | 2.77 | 1.12 |
| 244 | E | 2.29 | 1.00 |
| 246 | E | 2.97 | 1.54 |
| 257 | E | 2.17 | 1.24 |
| 266 | E | 3.37 | 1.31 |
| 270 | E | 3.55 | 1.03 |
| 272 | E | 3.55 | 1.03 |
| 273 | E | 3.21 | 1.00 |
| 288 | E | 3.41 | .93 |
| 289 | E | 3.46 | .97 |
| 298 | E | 2.54 | 1.35 |
| 300 | E | 3.67 | .80 |
| | | | |
| 2 | N | 3.01 | 1.12 |
| 30 | N | 2.29 | .80 |
| 32 | N | 3.13 | 1.43 |
| 35 | N | 2.71 | 1.16 |
| 36 | N | 2.54 | 1.64 |
| 40 | N | 2.52 | 1.48 |
| 43 | N | 2.74 | 1.16 |
| 44 | N | 3.13 | 1.56 |
| 48 | N | 2.84 | .88 |

| | | | |
|---|---|---|---|
| 49 | N | 2.55 | 1.25 |
| 53 | N | 2.28 | 1.44 |
| 60 | N | 3.04 | 1.50 |
| 61 | N | 2.66 | 1.19 |
| 62 | N | 2.56 | 1.23 |
| 64 | N | 2.81 | 1.69 |
| 72 | N | 2.65 | 1.31 |
| 79 | N | 2.29 | 1.26 |
| 82 | N | 2.95 | 1.24 |
| 91 | N | 2.80 | 1.28 |
| 99 | N | 2.18 | 1.03 |
| 100 | N | 2.62 | 1.66 |
| 107 | N | 3.03 | 1.23 |
| 117 | N | 2.31 | .82 |
| 119 | N | 1.98 | 1.175 |
| 122 | N | 2.99 | 1.50 |
| 123 | N | 3.11 | 1.17 |
| 133 | N | 3.09 | 1.18 |
| 134 | N | 2.98 | 1.09 |
| 135 | N | 2.83 | 1.27 |
| 137 | N | 3.20 | 1.40 |
| 138 | N | 3.33 | 1.47 |
| 153 | N | 3.56 | 1.22 |
| 156 | N | 3.19 | 1.46 |
| 161 | N | 2.34 | 1.02 |
| 166 | N | 2.98 | 1.41 |
| 169 | N | 2.99 | 1.25 |
| 173 | N | 2.77 | 1.07 |
| 182 | N | 2.28 | .81 |
| 184 | N | 2.21 | 1.12 |
| 188 | N | 2.97 | 1.48 |
| 192 | N | 2.81 | 1.45 |
| 203 | N | 2.75 | 1.40 |
| 206 | N | 3.05 | 1.24 |
| 216 | N | 2.52 | 1.46 |
| 217 | N | 2.82 | 1.67 |
| 218 | N | 2.90 | 1.56 |
| 222 | N | 2.64 | .98 |
| 226 | N | 3.74 | 1.22 |
| 233 | N | 3.31 | .89 |
| 241 | N | 2.68 | 1.65 |
| 243 | N | 3.06 | 1.49 |
| 247 | N | 3.34 | 1.24 |
| 254 | N | 2.99 | 1.19 |
| 255 | N | 2.63 | .84 |
| 263 | N | 2.68 | 1.18 |

| | | | |
|---|---|---|---|
| 265 | N | 2.24 | 1.19 |
| 278 | N | 2.45 | 1.17 |
| 281 | N | 3.04 | 1.38 |
| 287 | N | 3.65 | 1.05 |
| 292 | N | 2.29 | 1.05 |
| | | | |
| 3 | O | 3.75 | .89 |
| 6 | O | 2.95 | 2.52 |
| 16 | O | 2.09 | 1.08 |
| 17 | O | 3.51 | 1.47 |
| 19 | O | 3.33 | 1.76 |
| 28 | O | 4.34 | .88 |
| 29 | O | 3.31 | 1.54 |
| 51 | O | 4.30 | .80 |
| 54 | O | 3.81 | .88 |
| 58 | O | 3.68 | 1.19 |
| 65 | O | 3.46 | 1.12 |
| 66 | O | 3.47 | 1.48 |
| 69 | O | 3.24 | 1.30 |
| 75 | O | 3.84 | 1.38 |
| 77 | O | 3.53 | 1.06 |
| 78 | O | 2.88 | 1.01 |
| 85 | O | 3.76 | .83 |
| 86 | O | 3.27 | 1.60 |
| 87 | O | 2.29 | .97 |
| 89 | O | 3.09 | 1.17 |
| 90 | O | 3.82 | .96 |
| 94 | O | 3.72 | 1.00 |
| 103 | O | 3.91 | 1.11 |
| 104 | O | 3.67 | 1.01 |
| 113 | O | 3.49 | .80 |
| 125 | O | 3.23 | 1.20 |
| 140 | O | 3.42 | 1.31 |
| 141 | O | 2.93 | 1.32 |
| 142 | O | 3.51 | 1.16 |
| 144 | O | 3.32 | 1.72 |
| 145 | O | 3.30 | 1.30 |
| 160 | O | 4.02 | .97 |
| 170 | O | 3.15 | 1.82 |
| 174 | O | 3.37 | 1.24 |
| 175 | O | 2.96 | .92 |
| 180 | O | 3.37 | 1.24 |
| 194 | O | 3.19 | 1.63 |
| 195 | O | 2.68 | 1.30 |
| 202 | O | 3.12 | 1.19 |
| 208 | O | 2.65 | 1.59 |

| | | | |
|---|---|---|---|
| 211 | O | 3.70 | 1.22 |
| 215 | O | 3.49 | 1.49 |
| 221 | O | 2.57 | 1.25 |
| 223 | O | 4.03 | .63 |
| 224 | O | 3.14 | 1.44 |
| 228 | O | 3.58 | 1.82 |
| 235 | O | 3.43 | 1.37 |
| 239 | O | 4.00 | .99 |
| 252 | O | 3.42 | 1.11 |
| 258 | O | 3.44 | 1.21 |
| 260 | O | 3.39 | 1.40 |
| 264 | O | 3.35 | 1.19 |
| 268 | O | 3.68 | 1.14 |
| 275 | O | 2.99 | .94 |
| 277 | O | 3.76 | 1.23 |
| 282 | O | 3.15 | 1.11 |
| 283 | O | 3.93 | .67 |
| 285 | O | 4.46 | .65 |
| 291 | O | 3.07 | 1.42 |
| 293 | O | 3.66 | 1.53 |
| | | | |
| 4 | A | 3.95 | .935 |
| 5 | A | 3.36 | 1.166 |
| 8 | A | 4.00 | .835 |
| 9 | A | 3.11 | 1.176 |
| 13 | A | 3.22 | .89 |
| 14 | A | 3.97 | 1.18 |
| 15 | A | 3.28 | 1.09 |
| 20 | A | 3.66 | 1.25 |
| 23 | A | 4.27 | .75 |
| 25 | A | 2.53 | 1.21 |
| 26 | A | 3.42 | 1.06 |
| 27 | A | 3.17 | 1.39 |
| 31 | A | 3.21 | 1.11 |
| 37 | A | 3.74 | 1.25 |
| 42 | A | 3.23 | 1.10 |
| 68 | A | 3.45 | 1.36 |
| 70 | A | 3.99 | .87 |
| 71 | A | 3.90 | .87 |
| 73 | A | 3.07 | 1.36 |
| 83 | A | 3.71 | 1.12 |
| 84 | A | 3.64 | 1.24 |
| 93 | A | 3.29 | .83 |
| 109 | A | 3.24 | 1.32 |
| 114 | A | 2.85 | .83 |
| 118 | A | 3.55 | 1.00 |

| | | | |
|---|---|---|---|
| 126 | A | 3.79 | .85 |
| 127 | A | 3.84 | 1.25 |
| 129 | A | 3.31 | 1.22 |
| 130 | A | 3.41 | 1.83 |
| 131 | A | 2.81 | 1.61 |
| 132 | A | 2.95 | 1.04 |
| 143 | A | 3.14 | .99 |
| 147 | A | 3.25 | 1.85 |
| 149 | A | 3.07 | .83 |
| 154 | A | 3.78 | 1.24 |
| 165 | A | 3.77 | 1.21 |
| 168 | A | 2.60 | .96 |
| 179 | A | 3.25 | 1.45 |
| 189 | A | 3.84 | 1.34 |
| 191 | A | 2.80 | 1.15 |
| 199 | A | 3.81 | .819 |
| 225 | A | 3.02 | 1.66 |
| 227 | A | 3.51 | 1.40 |
| 232 | A | 3.73 | 1.06 |
| 236 | A | 2.83 | 1.00 |
| 238 | A | 3.21 | 1.30 |
| 240 | A | 3.51 | 1.45 |
| 242 | A | 3.70 | .98 |
| 249 | A | 3.37 | 1.05 |
| 250 | A | 2.81 | 1.47 |
| 251 | A | 2.63 | 1.09 |
| 259 | A | 3.55 | 1.05 |
| 261 | A | 3.04 | 1.39 |
| 262 | A | 2.76 | 1.34 |
| 269 | A | 4.01 | 1.04 |
| 276 | A | 2.75 | 1.08 |
| 279 | A | 3.57 | 1.17 |
| 280 | A | 3.90 | 1.11 |
| 284 | A | 3.68 | 1.14 |
| 297 | A | 3.65 | 1.01 |

Note. C=Conscientiousness, E=Extraversion, N= Neuroticism, O=Openness, A= Agreeableness

Table 5

*MANOVA Means and Partial Eta Squared for Each IPIP Item*

| Item | Factor | Men | Women | *p* | Partial eta squared |
|---|---|---|---|---|---|
| 1 | C | 3.91 | 3.81 | .365 | .003 |
| **2** | **N** | **2.73** | **3.15** | **.002** | **.038** |
| 3 | O | 3.84 | 3.71 | .315 | .004 |
| 4 | A | 3.92 | 3.94 | .878 | .000 |
| 5 | A | 3.46 | 3.26 | .144 | .009 |
| 6 | O | 2.84 | 2.94 | .617 | .001 |
| 7 | C | 3.45 | 3.38 | .694 | .001 |
| 8 | A | 4.11 | 3.94 | .144 | .009 |
| **9** | **A** | **3.26** | **2.99** | **.047** | **.016** |
| 10 | E | 3.96 | 3.94 | .895 | .000 |
| 11 | E | 3.47 | 3.70 | .171 | .008 |
| 12 | C | 3.24 | 3.21 | .843 | .000 |
| 13 | A | 3.28 | 3.20 | .487 | .002 |
| 14 | A | 4.00 | 3.99 | .963 | .000 |
| 15 | A | 3.32 | 3.28 | .774 | .000 |
| 16 | O | 2.09 | 2.09 | .981 | .000 |
| **17** | **O** | **3.13** | **3.73** | **.000** | **.058** |
| 18 | E | 3.27 | 3.23 | .824 | .000 |
| 19 | O | 3.32 | 3.41 | .619 | .001 |
| 20 | A | 3.68 | 3.70 | .897 | .000 |
| 21 | E | 3.76 | 3.71 | .715 | .001 |
| 22 | C | 4.11 | 4.21 | .433 | .002 |
| 23 | A | 4.33 | 4.28 | .678 | .001 |
| 24 | E | 3.51 | 3.42 | .509 | .002 |
| 25 | A | 2.40 | 2.62 | .119 | .010 |
| 26 | A | 3.45 | 3.43 | .881 | .000 |
| 27 | A | 3.12 | 3.17 | .702 | .001 |
| 28 | O | 4.30 | 4.34 | .730 | .000 |
| 29 | O | 3.22 | 3.34 | .447 | .002 |
| 30 | N | 2.18 | 2.34 | .188 | .007 |
| 31 | A | 3.28 | 3.16 | .367 | .003 |
| 32 | N | 3.18 | 3.12 | .658 | .001 |
| **33** | **C** | **3.76** | **4.03** | **.042** | **.017** |
| 34 | C | 3.59 | 3.63 | .698 | .001 |
| 35 | N | 2.62 | 2.79 | .244 | .005 |
| 36 | N | 2.36 | 2.63 | .095 | .011 |
| 37 | A | 3.78 | 3.72 | .665 | .001 |
| 38 | E | 3.94 | 3.84 | .425 | .003 |
| **39** | **E** | **3.37** | **3.03** | **.039** | **.017** |
| 40 | N | 2.37 | 2.60 | .140 | .009 |
| 41 | E | 4.14 | 3.91 | .099 | .011 |

| | | | | | |
|---|---|---|---|---|---|
| 42 | A | 3.29 | 3.17 | .374 | .003 |
| 43 | N | 2.65 | 2.71 | .685 | .001 |
| 44 | N | 3.00 | 3.24 | .138 | .009 |
| 45 | C | 4.04 | 4.00 | .704 | .001 |
| 46 | C | 3.60 | 3.51 | .464 | .002 |
| 47 | C | 4.11 | 4.07 | .789 | .000 |
| 48 | N | 2.71 | 2.92 | .097 | .011 |
| 49 | N | 2.49 | 2.57 | .560 | .001 |
| 50 | C | 3.77 | 3.72 | .685 | .001 |
| 51 | O | 4.30 | 4.31 | .968 | .000 |
| 52 | E | 3.83 | 3.75 | .568 | .001 |
| 53 | N | 2.19 | 2.30 | .484 | .002 |
| 54 | O | 3.91 | 3.81 | .388 | .003 |
| 55 | E | 3.67 | 3.59 | .601 | .001 |
| **56** | **C** | **3.92** | **3.61** | **.026** | **.020** |
| 57 | C | 3.94 | 3.88 | .659 | .001 |
| **58** | **O** | **3.49** | **3.83** | **.020** | **.022** |
| 59 | C | 3.47 | 3.46 | .929 | .000 |
| 60 | N | 2.88 | 3.08 | .191 | .007 |
| **61** | **N** | **2.37** | **2.81** | **.001** | **.041** |
| **62** | **N** | **2.33** | **2.67** | **.015** | **.024** |
| 63 | E | 3.71 | 3.77 | .689 | .001 |
| 64 | N | 2.61 | 2.93 | .061 | .014 |
| 65 | O | 3.44 | 3.46 | .898 | .000 |
| 66 | O | 3.42 | 3.53 | .489 | .002 |
| 67 | E | 3.50 | 3.45 | .704 | .001 |
| **68** | **A** | **3.25** | **3.55** | **.049** | **.016** |
| 69 | O | 3.23 | 3.24 | .947 | .000 |
| 70 | A | 3.89 | 4.03 | .281 | .005 |
| 71 | A | 3.92 | 3.89 | .827 | .000 |
| 72 | N | 2.57 | 2.64 | .649 | .001 |
| 73 | A | 3.23 | 2.98 | .096 | .011 |
| 74 | E | 2.91 | 2.85 | .711 | .001 |
| 75 | O | 3.76 | 3.99 | .133 | .009 |
| 76 | E | 3.18 | 3.08 | .531 | .002 |
| 77 | O | 3.58 | 3.48 | .475 | .002 |
| 78 | O | 2.91 | 2.83 | .515 | .002 |
| 79 | N | 2.30 | 2.22 | .599 | .001 |
| 80 | C | 3.43 | 3.56 | .391 | .003 |
| 81 | E | 3.73 | 3.62 | .438 | .002 |
| **82** | **N** | **2.68** | **3.08** | **.007** | **.029** |
| 83 | A | 3.58 | 3.80 | .100 | .011 |
| 84 | A | 3.65 | 3.64 | .924 | .000 |
| **85** | **O** | **3.92** | **3.69** | **.050** | **.015** |
| 86 | O | 3.29 | 3.32 | .836 | .000 |
| 87 | O | 2.37 | 2.25 | .358 | .003 |

| | | | | | |
|---|---|---|---|---|---|
| 88 | C | 3.32 | 3.29 | .827 | .000 |
| 89 | O | 3.08 | 3.11 | .818 | .000 |
| 90 | O | 3.88 | 3.80 | .527 | .002 |
| 91 | N | 2.67 | 2.92 | .087 | .012 |
| 92 | E | 3.41 | 3.32 | .532 | .002 |
| 93 | A | 3.36 | 3.26 | .428 | .003 |
| 94 | O | 3.77 | 3.69 | .538 | .002 |
| 95 | E | 3.87 | 3.68 | .163 | .008 |
| 96 | C | 3.70 | 3.70 | .999 | .000 |
| 97 | C | 3.16 | 3.28 | .451 | .002 |
| 98 | C | 3.97 | 3.94 | .795 | .000 |
| 99 | N | 2.17 | 2.15 | .848 | .000 |
| 100 | N | 2.44 | 2.67 | .166 | .008 |
| **101** | **C** | **2.70** | **3.00** | **.042** | **.017** |
| 102 | C | 3.03 | 3.07 | .803 | .000 |
| 103 | O | 3.96 | 3.96 | .971 | .000 |
| 104 | O | 3.57 | 3.73 | .211 | .006 |
| 105 | E | 2.15 | 2.05 | .410 | .003 |
| 106 | C | 3.61 | 3.56 | .685 | .001 |
| 107 | N | 3.06 | 3.00 | .672 | .001 |
| 108 | C | 3.10 | 2.98 | .437 | .002 |
| 109 | A | 3.09 | 3.34 | .099 | .011 |
| 110 | E | 3.34 | 3.39 | .678 | .001 |
| 111 | E | 3.41 | 3.63 | .114 | .010 |
| 112 | E | 3.46 | 3.43 | .842 | .000 |
| 113 | O | 3.51 | 3.50 | .985 | .000 |
| 114 | A | 2.85 | 2.79 | .578 | .001 |
| 115 | C | 3.75 | 3.85 | .471 | .002 |
| 116 | C | 3.31 | 3.28 | .776 | .000 |
| 117 | N | 2.28 | 2.34 | .567 | .001 |
| 118 | A | 3.52 | 3.60 | .516 | .002 |
| 119 | N | 2.08 | 1.86 | .112 | .010 |
| **120** | **C** | **3.86** | **4.10** | **.045** | **.016** |
| 121 | C | 3.49 | 3.42 | .640 | .001 |
| 122 | N | 3.00 | 3.00 | 1.000 | .000 |
| 123 | N | 3.00 | 3.14 | .318 | .004 |
| 124 | E | 3.37 | 3.55 | .140 | .009 |
| **125** | **O** | **3.00** | **3.34** | **.020** | **.022** |
| 126 | A | 3.87 | 3.75 | .317 | .004 |
| 127 | A | 3.77 | 3.89 | .402 | .003 |
| 128 | C | 3.85 | 3.78 | .513 | .002 |
| 129 | A | 3.36 | 3.32 | .802 | .000 |
| 130 | A | 3.38 | 3.52 | .438 | .002 |
| 131 | A | 2.89 | 2.82 | .647 | .001 |
| 132 | A | 3.04 | 2.89 | .248 | .005 |
| **133** | **N** | **2.91** | **3.19** | **.047** | **.016** |

| | | | | | |
|---|---|---|---|---|---|
| 134 | N | 2.84 | 3.01 | .197 | .007 |
| **135** | **N** | **2.57** | **2.94** | **.009** | **.027** |
| 136 | E | 3.16 | 3.30 | .410 | .003 |
| 137 | N | 3.06 | 3.26 | .186 | .007 |
| 138 | N | 3.23 | 3.37 | .396 | .003 |
| 139 | C | 3.32 | 3.32 | .988 | .000 |
| 140 | O | 3.43 | 3.38 | .754 | .000 |
| 141 | O | 3.09 | 2.86 | .126 | .009 |
| 142 | O | 3.61 | 3.49 | .375 | .003 |
| 143 | A | 3.23 | 3.09 | .275 | .005 |
| 144 | O | 3.34 | 3.37 | .881 | .000 |
| 145 | O | 3.39 | 3.27 | .428 | .003 |
| 146 | E | 3.77 | 3.61 | .337 | .004 |
| 147 | A | 3.12 | 3.35 | .191 | .007 |
| 148 | E | 3.36 | 3.13 | .146 | .009 |
| 149 | A | 2.96 | 3.12 | .161 | .008 |
| 150 | E | 3.93 | 3.79 | .401 | .003 |
| 151 | C | 3.98 | 3.99 | .948 | .000 |
| 152 | E | 2.40 | 2.52 | .340 | .004 |
| 153 | N | 3.43 | 3.63 | .153 | .008 |
| 154 | A | 3.86 | 3.74 | .401 | .003 |
| 155 | E | 3.63 | 3.45 | .170 | .008 |
| 156 | N | 3.15 | 3.17 | .891 | .000 |
| 157 | C | 3.90 | 3.81 | .392 | .003 |
| 158 | C | 3.48 | 3.46 | .825 | .000 |
| 159 | C | 3.85 | 3.84 | .926 | .000 |
| 160 | O | 4.10 | 4.00 | .443 | .002 |
| **161** | **N** | **2.13** | **2.45** | **.012** | **.025** |
| 162 | E | 3.79 | 3.75 | .777 | .000 |
| 163 | C | 3.45 | 3.28 | .307 | .004 |
| 164 | E | 3.31 | 3.36 | .701 | .001 |
| 165 | A | 3.84 | 3.77 | .632 | .001 |
| 166 | N | 2.81 | 3.06 | .098 | .011 |
| 167 | C | 3.58 | 3.47 | .438 | .002 |
| 168 | A | 2.61 | 2.58 | .817 | .000 |
| 169 | N | 2.87 | 3.00 | .361 | .003 |
| 170 | O | 3.14 | 3.22 | .666 | .001 |
| 171 | E | 3.67 | 3.47 | .234 | .006 |
| **172** | **E** | **3.37** | **3.05** | **.056** | **.015** |
| 173 | N | 2.78 | 2.75 | .835 | .000 |
| 174 | O | 3.38 | 3.36 | .892 | .000 |
| 175 | O | 2.82 | 3.04 | .080 | .012 |
| 176 | C | 3.83 | 3.80 | .834 | .000 |
| 177 | C | 4.00 | 3.78 | .082 | .012 |
| 178 | E | 3.07 | 2.99 | .595 | .001 |
| **179** | **A** | **3.03** | **3.37** | **.029** | **.019** |

| | | | | | |
|---|---|---|---|---|---|
| **180** | **O** | **3.17** | **3.46** | **.044** | **.016** |
| 181 | E | 3.74 | 3.71 | .768 | .000 |
| 182 | N | 2.18 | 2.30 | .316 | .004 |
| 183 | C | 3.49 | 3.68 | .181 | .007 |
| 184 | N | 2.23 | 2.20 | .803 | .000 |
| 185 | E | 3.32 | 3.41 | .560 | .001 |
| 186 | C | 3.61 | 3.65 | .776 | .000 |
| 187 | E | 3.39 | 3.34 | .718 | .001 |
| **188** | **N** | **2.67** | **3.04** | **.019** | **.022** |
| 189 | A | 3.75 | 3.93 | .227 | .006 |
| 190 | E | 3.23 | 3.23 | .991 | .000 |
| 191 | A | 2.68 | 2.81 | .379 | .003 |
| 192 | N | 2.67 | 2.84 | .293 | .004 |
| 193 | E | 3.44 | 3.37 | .755 | .000 |
| 194 | O | 3.17 | 3.23 | .730 | .000 |
| 195 | O | 2.67 | 2.69 | .887 | .000 |
| 196 | E | 3.48 | 3.48 | .970 | .000 |
| 197 | E | 2.32 | 2.49 | .183 | .007 |
| 198 | E | 3.48 | 3.65 | .229 | .006 |
| 199 | A | 3.81 | 3.82 | .940 | .000 |
| 200 | E | 3.85 | 3.71 | .355 | .003 |
| 201 | C | 3.73 | 3.59 | .266 | .005 |
| 202 | O | 3.08 | 3.18 | .471 | .002 |
| 203 | N | 2.67 | 2.74 | .683 | .001 |
| 204 | C | 3.24 | 3.23 | .966 | .005 |
| 205 | E | 3.49 | 3.40 | .437 | .002 |
| **206** | **N** | **2.85** | **3.15** | **.036** | **.017** |
| 207 | C | 3.66 | 3.81 | .252 | .005 |
| 208 | O | 2.78 | 2.53 | .122 | .010 |
| **209** | **E** | **3.79** | **3.48** | **.020** | **.022** |
| 210 | C | 4.08 | 3.92 | .118 | .010 |
| 211 | O | 3.77 | 3.68 | .494 | .002 |
| 212 | E | 3.99 | 3.83 | .165 | .008 |
| 213 | C | 3.14 | 3.07 | .626 | .001 |
| 214 | C | 3.87 | 3.82 | .710 | .001 |
| 215 | O | 3.47 | 3.52 | .759 | .001 |
| 216 | N | 2.47 | 2.51 | .799 | .000 |
| 217 | N | 2.81 | 2.83 | .876 | .000 |
| 218 | N | 2.87 | 2.92 | .765 | .000 |
| 219 | C | 2.92 | 3.06 | .448 | .002 |
| 220 | E | 4.22 | 4.08 | .184 | .007 |
| 221 | O | 2.43 | 2.67 | .104 | .011 |
| 222 | N | 2.60 | 2.66 | .667 | .001 |
| 223 | O | 3.96 | 4.10 | .182 | .007 |
| 224 | O | 3.27 | 3.14 | .412 | .003 |
| 225 | A | 3.00 | 3.01 | .939 | .000 |

| | | | | | |
|---|---|---|---|---|---|
| 226 | N | 3.67 | 3.82 | .304 | .004 |
| 227 | A | 3.46 | 3.54 | .642 | .001 |
| 228 | O | 3.49 | 3.65 | .378 | .003 |
| 229 | E | 3.26 | 3.23 | .881 | .000 |
| 230 | E | 2.41 | 2.33 | .540 | .002 |
| 231 | C | 3.89 | 3.78 | .366 | .003 |
| 232 | A | 3.89 | 3.66 | .078 | .012 |
| 233 | N | 3.17 | 3.37 | .112 | .010 |
| 234 | C | 3.04 | 3.15 | .464 | .002 |
| 235 | O | 3.44 | 3.46 | .907 | .000 |
| **236** | **A** | **2.57** | **2.99** | **.001** | **.043** |
| 237 | E | 2.82 | 2.73 | .514 | .002 |
| 238 | A | 3.33 | 3.15 | .247 | .005 |
| 239 | O | 4.10 | 3.99 | .398 | .003 |
| 240 | A | 3.48 | 3.51 | .870 | .000 |
| 241 | N | 2.46 | 2.76 | .068 | .013 |
| 242 | A | 3.60 | 3.75 | .244 | .005 |
| 243 | N | 2.84 | 3.12 | .069 | .013 |
| 244 | E | 2.37 | 2.23 | .260 | .005 |
| 245 | C | 3.17 | 3.26 | .523 | .002 |
| 246 | E | 3.05 | 2.89 | .315 | .004 |
| 247 | N | 3.28 | 3.34 | .647 | .001 |
| 248 | C | 3.33 | 3.44 | .433 | .002 |
| 249 | A | 3.28 | 3.43 | .262 | .005 |
| 250 | A | 2.80 | 2.84 | .806 | .000 |
| 251 | A | 2.76 | 2.52 | .072 | .013 |
| 252 | O | 3.55 | 3.37 | .199 | .007 |
| 253 | C | 4.05 | 3.99 | .625 | .001 |
| 254 | N | 2.86 | 3.03 | .212 | .006 |
| 255 | N | 2.59 | 2.66 | .578 | .001 |
| 256 | C | 3.25 | 3.10 | .300 | .004 |
| 257 | E | 2.27 | 2.06 | .167 | .008 |
| 258 | O | 3.41 | 3.43 | .901 | .000 |
| **259** | **A** | **3.68** | **3.43** | **.059** | **.014** |
| 260 | O | 3.34 | 3.41 | .655 | .001 |
| 261 | A | 3.03 | 3.03 | .970 | .000 |
| **262** | **A** | **2.58** | **2.88** | **.043** | **.016** |
| 263 | N | 2.61 | 2.71 | .493 | .002 |
| 264 | O | 3.29 | 3.39 | .500 | .002 |
| 265 | N | 2.32 | 2.09 | .104 | .010 |
| 266 | E | 3.40 | 3.40 | .977 | .000 |
| 267 | C | 3.92 | 3.96 | .803 | .000 |
| **268** | **O** | **3.51** | **3.78** | **.053** | **.015** |
| 269 | A | 4.03 | 4.01 | .848 | .000 |
| 270 | E | 3.61 | 3.50 | .468 | .002 |
| **271** | **C** | **3.83** | **4.06** | **.048** | **.016** |

| | | | | | |
|---|---|---|---|---|---|
| 272 | E | 3.44 | 3.62 | .168 | .008 |
| 273 | E | 3.19 | 3.19 | .974 | .000 |
| 274 | C | 3.38 | 3.40 | .828 | .000 |
| 275 | O | 2.95 | 3.01 | .600 | .001 |
| 276 | A | 2.86 | 2.70 | .222 | .006 |
| 277 | O | 3.67 | 3.83 | .267 | .005 |
| 278 | N | 2.52 | 2.37 | .292 | .004 |
| **279** | **A** | **3.36** | **3.74** | **.007** | **.029** |
| 280 | A | 3.87 | 3.93 | .677 | .001 |
| 281 | N | 2.90 | 3.11 | .170 | .008 |
| 282 | O | 3.16 | 3.17 | .975 | .000 |
| 283 | O | 3.92 | 3.95 | .821 | .000 |
| 284 | A | 3.72 | 3.67 | .745 | .000 |
| **285** | **O** | **4.29** | **4.54** | **.016** | **.023** |
| 286 | C | 3.39 | 3.20 | .155 | .008 |
| 287 | N | 3.61 | 3.66 | .728 | .000 |
| 288 | E | 3.46 | 3.43 | .840 | .000 |
| 289 | E | 3.57 | 3.41 | .188 | .007 |
| 290 | C | 2.39 | 2.62 | .115 | .010 |
| 291 | O | 3.12 | 3.06 | .698 | .001 |
| 292 | N | 2.12 | 2.33 | .113 | .010 |
| **293** | **O** | **3.26** | **3.91** | **.000** | **.064** |
| 294 | C | 3.05 | 3.13 | .642 | .001 |
| 295 | C | 3.12 | 3.13 | .915 | .000 |
| 296 | C | 3.59 | 3.65 | .602 | .001 |
| 297 | A | 3.60 | 3.72 | .332 | .004 |
| 298 | E | 2.60 | 2.46 | .358 | .003 |
| 299 | C | 3.18 | 2.99 | .179 | .007 |
| 300 | E | 3.67 | 3.65 | .832 | .000 |

Note. C=Conscientiousness, E=Extraversion, N= Neuroticism, O=Openness, A= Agreeableness

Table 6

*Eigenvalues from the Principal Axis Factor Analysis of the Observed Data*

| Factor | 1st eigenvalue | % Variance | 2nd eigenvalue | % Variance |
|---|---|---|---|---|
| Conscientiousness | 11.863 | 19.771 | 3.962 | 6.604 |
| Extraversion | 12.479 | 20.798 | 3.329 | 5.548 |
| Neuroticism | 14.197 | 23.661 | 3.909 | 6.515 |
| Openness | 8.891 | 14.819 | 3.580 | 5.967 |
| Agreeableness | 8.591 | 14.318 | 4.056 | 6.759 |

Table 7

*Eigenvalues from the Principle Analysis Factor of the Observed Data and the Eigenvalues from the Parallel Analysis*

| | Component | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Observed Data | 36.227 | 17.044 | 12.053 | 9.569 | 7.300 |
| Parallel Analysis | 4.096 | 3.986 | 3.889 | 3.807 | 3.734 |

Table 8

*Rotated Loadings of the IPIP Factor Scores*

| | Component | | | | |
|---|---|---|---|---|---|
| Sub-scale | 1 | 2 | 3 | 4 | 5 |
| $C_1$ | .792 | | | | |
| $C_2$ | .318 | | | **.423** | |
| $C_3$ | .548 | .511 | | | |
| $C_4$ | .729 | | | **.391** | |
| $C_5$ | .719 | | | **.310** | |
| $C_6$ | .435 | .469 | | | |
| $E_1$ | **.715** | | | | -.336 |
| $E_2$ | **.566** | -.374 | | | |
| $E_3$ | **.636** | -.460 | | | |
| $E_4$ | **.475** | | | | |
| $E_5$ | | -.624 | | | |
| $E_6$ | **.649** | | | | |
| $N_1$ | -.581 | | **.577** | .332 | |
| $N_2$ | -.454 | | **.398** | .330 | |
| $N_3$ | -.661 | | **.424** | | |
| $N_4$ | -.639 | .336 | **.312** | | |
| $N_5$ | | | **.471** | | |
| $N_6$ | -.641 | | **.538** | | |
| $O_1$ | | | .430 | | |
| $O_2$ | .490 | | .438 | | |
| $O_3$ | | | .659 | | |
| $O_4$ | .547 | | | | |
| $O_5$ | .661 | | | | **.448** |
| $O_6$ | | | .317 | | **.345** |
| $A_1$ | .349 | | | -.404 | |
| $A_2$ | .375 | **.661** | | | |
| $A_3$ | .581 | **.376** | .466 | | |
| $A_4$ | | **.552** | | | |
| $A_5$ | | **.552** | | | |
| $A_6$ | | **.441** | .526 | | |

Note. Only loading >.30 are displayed. Loadings on the five personality are reported in bold.

Subscripts reflect the number subscale for each factor. C=Conscientiousness, E=Extraversion,

N= Neuroticism, O=Openness, A= Agreeableness

Table 9

*Fit Statistics for Samejima's Graded Response Model*

Frequency Table of Chi-Square/DF Ratios for Agreeableness

**Frequency Table of Chi-Square/DF Ratios**

| | <1 | 1<2 | 2<3 | 3<4 | 4<5 | 5<7 | >7 | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| **Singles** | 60 | 0 | 0 | 0 | 0 | 0 | 0 | 0.025 | 0.032 |
| **Doubles** | 25 | 30 | 5 | 0 | 0 | 0 | 0 | 1.165 | 0.525 |
| **Triples** | 4 | 16 | 0 | 0 | 0 | 0 | 0 | 1.31 | 0.355 |

Frequency Table of Chi-Square/DF Ratios for Conscientiousness

**Frequency Table of Chi-Square/DF Ratios**

| | <1 | 1<2 | 2<3 | 3<4 | 4<5 | 5<7 | >7 | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| **Singles** | 60 | 0 | 0 | 0 | 0 | 0 | 0 | 0.067 | 0.058 |
| **Doubles** | 30 | 27 | 3 | 0 | 0 | 0 | 0 | 1.054 | 0.467 |
| **Triples** | 7 | 13 | 0 | 0 | 0 | 0 | 0 | 1.171 | 0.312 |

Frequency Table of Chi-Square/ DF Ratios for Extraversion

**Frequency Table of Chi-Square/DF Ratios**

| | <1 | 1<2 | 2<3 | 3<4 | 4<5 | 5<7 | >7 | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| **Singles** | 60 | 0 | 0 | 0 | 0 | 0 | 0 | 0.032 | 0.034 |
| **Doubles** | 35 | 21 | 2 | 0 | 1 | 1 | 0 | 1.147 | 1.018 |
| **Triples** | 3 | 15 | 1 | 1 | 0 | 0 | 0 | 1.308 | 0.655 |

Frequency Table of Chi-Square/ DF Ratios for Neuroticism

**Frequency Table of Chi-Square/DF Ratios**

| | <1 | 1<2 | 2<3 | 3<4 | 4<5 | 5<7 | >7 | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| **Singles** | 60 | 0 | 0 | 0 | 0 | 0 | 0 | 0.024 | 0.02 |
| **Doubles** | 36 | 23 | 1 | 0 | 0 | 0 | 0 | 0.927 | 0.453 |
| **Triples** | 9 | 11 | 0 | 0 | 0 | 0 | 0 | 1.108 | 0.395 |

Frequency Table of Chi-Square/ DF Ratios for Openness

**Frequency Table of Chi-Square/DF Ratios**

| | <1 | 1<2 | 2<3 | 3<4 | 4<5 | 5<7 | >7 | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| **Singles** | 60 | 0 | 0 | 0 | 0 | 0 | 0 | 0.049 | 0.052 |
| **Doubles** | 33 | 24 | 2 | 1 | 0 | 0 | 0 | 1.017 | 0.595 |
| **Triples** | 6 | 12 | 2 | 0 | 0 | 0 | 0 | 1.237 | 0.532 |

Table 10

*Results of the Differential Item Functioning Analysis using the Mantel Procedure on the IPIP Items*

| Item | Factor | $\chi^2_M$ |
|---|---|---|
| 1 | C | 9.217* |
| 2 | N | 6.794 |
| 3 | O | 1.241 |
| 4 | A | 0.063 |
| 5 | A | 2.735 |
| 6 | O | 1.635 |
| 7 | C | 2.887 |
| 8 | A | 2.239 |
| 9 | A | 5.341 |
| 10 | E | 1.002 |
| 11 | E | 3.991 |
| 12 | C | 0.282 |
| 13 | A | 0.265 |
| 14 | A | 1.753 |
| 15 | A | 4.277 |
| 16 | O | 0.387 |
| 17 | O | 15.086** |
| 18 | E | 5.565 |
| 19 | O | 3.523 |
| 20 | A | 2.444 |
| 21 | E | 3.233 |
| 22 | C | 5.021 |
| 23 | A | 9.367* |
| 24 | E | 1.141 |
| 25 | A | 4.283 |
| 26 | A | 1.842 |
| 27 | A | 0.947 |
| 28 | O | 3.068 |
| 29 | O | 6.151 |
| 30 | N | 4.911 |
| 31 | A | 5.617 |
| 32 | N | 5.408 |
| 33 | C | 10.340** |
| 34 | C | 6.581 |
| 35 | N | 3.853 |
| 36 | N | 1.498 |
| 37 | A | 6.095 |
| 38 | E | 1.525 |

| | | |
|---|---|---|
| 39 | E | 5.193 |
| 40 | N | 2.018 |
| 41 | E | 2.089 |
| 42 | A | 2.116 |
| 43 | N | 4.826 |
| 44 | N | 1.133 |
| 45 | C | 0.786 |
| 46 | C | 1.826 |
| 47 | C | 5.261 |
| 48 | N | 3.732 |
| 49 | N | 3.145 |
| 50 | C | 2.794 |
| 51 | O | 1.621 |
| 52 | E | 2.213 |
| 53 | N | 1.854 |
| 54 | O | 3.189 |
| 55 | E | 1.185 |
| 56 | C | 7.915* |
| 57 | C | 2.262 |
| 58 | O | 4.922 |
| 59 | C | 7.357* |
| 60 | N | 2.093 |
| 61 | N | 11.538** |
| 62 | N | 5.025 |
| 63 | E | 4.799 |
| 64 | N | 5.854 |
| 65 | O | 1.043 |
| 66 | O | 3.733 |
| 67 | E | 0.526 |
| 68 | A | 4.788 |
| 69 | O | 0.830 |
| 70 | A | 4.136 |
| 71 | A | 0.168 |
| 72 | N | 0.235 |
| 73 | A | 4.283 |
| 74 | E | 2.545 |
| 75 | O | 3.716 |
| 76 | E | 0.530 |
| 77 | O | 0.787 |
| 78 | O | 2.978 |
| 79 | N | 6.393 |
| 80 | C | 5.133 |
| 81 | E | 1.504 |
| 82 | N | 4.223 |
| 83 | A | 5.832 |
| 84 | A | 1.809 |

| | | |
|---|---|---|
| 85 | O | 4.420 |
| 86 | O | 2.021 |
| 87 | O | 1.115 |
| 88 | C | 1.400 |
| 89 | O | 2.327 |
| 90 | O | 3.181 |
| 91 | N | 1.222 |
| 92 | E | 1.131 |
| 93 | A | 1.599 |
| 94 | O | 3.404 |
| 95 | E | 5.161 |
| 96 | C | 0.502 |
| 97 | C | 0.281 |
| 98 | C | 4.836 |
| 99 | N | 0.614 |
| 100 | N | 2.406 |
| 101 | C | 5.026 |
| 102 | C | 2.051 |
| 103 | O | 0.622 |
| 104 | O | 6.254 |
| 105 | E | 0.986 |
| 106 | C | 0.171 |
| 107 | N | 2.125 |
| 108 | C | 2.473 |
| 109 | A | 2.128 |
| 110 | E | 8.129* |
| 111 | E | 3.319 |
| 112 | E | 1.433 |
| 113 | O | 1.057 |
| 114 | A | 4.329 |
| 115 | C | 2.292 |
| 116 | C | 3.485 |
| 117 | N | 1.523 |
| 118 | A | 1.397 |
| 119 | N | 5.578 |
| 120 | C | 5.641 |
| 121 | C | 3.186 |
| 122 | N | 4.463 |
| 123 | N | 2.657 |
| 124 | E | 4.600 |
| 125 | O | 7.432* |
| 126 | A | 4.216 |
| 127 | A | 1.426 |
| 128 | C | 1.203 |
| 129 | A | 3.272 |
| 130 | A | 1.091 |

| | | |
|---|---|---|
| 131 | A | 1.778 |
| 132 | A | 8.457* |
| 133 | N | 7.608* |
| 134 | N | 2.455 |
| 135 | N | 6.031 |
| 136 | E | 2.211 |
| 137 | N | 4.343 |
| 138 | N | 2.882 |
| 139 | C | 2.410 |
| 140 | O | 0.633 |
| 141 | O | 6.586 |
| 142 | O | 1.261 |
| 143 | A | 1.477 |
| 144 | O | 3.822 |
| 145 | O | 4.732 |
| 146 | E | 3.461 |
| 147 | A | 7.085* |
| 148 | E | 4.653 |
| 149 | A | 2.555 |
| 150 | E | 2.423 |
| 151 | C | 1.983 |
| 152 | E | 4.287 |
| 153 | N | 0.764 |
| 154 | A | 1.761 |
| 155 | E | 0.991 |
| 156 | N | 0.809 |
| 157 | C | 2.939 |
| 158 | C | 1.597 |
| 159 | C | 3.929 |
| 160 | O | 6.885 |
| 161 | N | 3.486 |
| 162 | E | 0.332 |
| 163 | C | 2.008 |
| 164 | E | 4.786 |
| 165 | A | 2.597 |
| 166 | N | 2.317 |
| 167 | C | 2.377 |
| 168 | A | 1.289 |
| 169 | N | 7.099 |
| 170 | O | 0.133 |
| 171 | E | 2.854 |
| 172 | E | 3.267 |
| 173 | N | 0.928 |
| 174 | O | 4.216 |
| 175 | O | 6.143 |
| 176 | C | 1.806 |

| | | |
|---|---|---|
| 177 | C | 3.335 |
| 178 | E | 0.771 |
| 179 | A | 3.805 |
| 180 | O | 3.027 |
| 181 | E | 5.713 |
| 182 | N | 0.305 |
| 183 | C | 4.267 |
| 184 | N | 7.596* |
| 185 | E | 2.907 |
| 186 | C | 2.190 |
| 187 | E | 0.429 |
| 188 | N | 5.057 |
| 189 | A | 6.126 |
| 190 | E | 4.014 |
| 191 | A | 1.631 |
| 192 | N | 5.939 |
| 193 | E | 2.540 |
| 194 | O | 2.588 |
| 195 | O | 1.749 |
| 196 | E | 3.456 |
| 197 | E | 3.074 |
| 198 | E | 2.619 |
| 199 | A | 0.412 |
| 200 | E | 0.537 |
| 201 | C | 9.688* |
| 202 | O | 2.674 |
| 203 | N | 0.758 |
| 204 | C | 2.779 |
| 205 | E | 7.034 |
| 206 | N | 4.010 |
| 207 | C | 2.626 |
| 208 | O | 8.875* |
| 209 | E | 7.293 |
| 210 | C | 3.630 |
| 211 | O | 1.588 |
| 212 | E | 9.691* |
| 213 | C | 0.339 |
| 214 | C | 3.854 |
| 215 | O | 1.904 |
| 216 | N | 4.221 |
| 217 | N | 4.321 |
| 218 | N | 1.480 |
| 219 | C | 2.798 |
| 220 | E | 7.232 |
| 221 | O | 4.971 |
| 222 | N | 5.143 |

| | | |
|---|---|---|
| 223 | O | 7.885* |
| 224 | O | 2.864 |
| 225 | A | 4.312 |
| 226 | N | 1.207 |
| 227 | A | 1.542 |
| 228 | O | 1.146 |
| 229 | E | 3.790 |
| 230 | E | 3.753 |
| 231 | C | 2.635 |
| 232 | A | 5.478 |
| 233 | N | 2.004 |
| 234 | C | 2.289 |
| 235 | O | 2.687 |
| 236 | A | 9.767* |
| 237 | E | 3.930 |
| 238 | A | 7.603* |
| 239 | O | 4.492 |
| 240 | A | 1.605 |
| 241 | N | 2.059 |
| 242 | A | 1.474 |
| 243 | N | 1.277 |
| 244 | E | 3.493 |
| 245 | C | 0.126 |
| 246 | E | .0624 |
| 247 | N | 1.436 |
| 248 | C | 1.318 |
| 249 | A | 2.882 |
| 250 | A | 2.385 |
| 251 | A | 7.564* |
| 252 | O | 5.394 |
| 253 | C | 0.652 |
| 254 | N | 1.688 |
| 255 | N | 1.102 |
| 256 | C | 4.505 |
| 257 | E | 6.510 |
| 258 | O | 0.953 |
| 259 | A | 8.319* |
| 260 | O | 0.429 |
| 261 | A | 0.585 |
| 262 | A | 0.191 |
| 263 | N | 1.408 |
| 264 | O | 2.320 |
| 265 | N | 12.317** |
| 266 | E | 0.478 |
| 267 | C | 2.973 |
| 268 | O | 6.124 |

| | | |
|---|---|---|
| 269 | A | 1.554 |
| 270 | E | 0.665 |
| 271 | C | 2.431 |
| 272 | E | 3.575 |
| 273 | E | 0.563 |
| 274 | C | 2.410 |
| 275 | O | 1.938 |
| 276 | A | 1.637 |
| 277 | O | 3.568 |
| 278 | N | 6.157 |
| 279 | A | 6.662 |
| 280 | A | 1.314 |
| 281 | N | 2.129 |
| 282 | O | 3.675 |
| 283 | O | 2.054 |
| 284 | A | 0.619 |
| 285 | O | 9.238* |
| 286 | C | 3.352 |
| 287 | N | 0.764 |
| 288 | E | 8.731* |
| 289 | E | 4.844 |
| 290 | C | 4.623 |
| 291 | O | 0.376 |
| 292 | N | 2.495 |
| 293 | O | 16.144** |
| 294 | C | 7.358* |
| 295 | C | 3.647 |
| 296 | C | 1.434 |
| 297 | A | 0.151 |
| 298 | E | 2.613 |
| 299 | C | 5.078 |
| 300 | E | 2.754 |

Note. C=Conscientiousness, E=Extraversion, N= Neuroticism, O=Openness, A= Agreeableness.

*$p < .05$, *** $p < .01$, $\chi^2_M$ = Mantel Chi-Square.

Table 11

*Linking Coefficients from the Mean-Sigma Method*

| Factor | α | β |
|---|---|---|
| E | 1.110 | 0.177 |
| O | 1.144 | 0.026 |
| N | 0.989 | -0.288 |
| C | 1.031 | -0.023 |
| A | 1.129 | -0.041 |

Note. C=Conscientiousness, E=Extraversion, N= Neuroticism, O=Openness, A= Agreeableness.

Table 12

*Results of the Differential Item Functioning Analysis using the R program adaptation of DFIT Procedure*

| Item | Factor | NCDIF |
|---|---|---|
| 1 | C | .0261 |
| 2 | N | .0529 |
| 3 | O | .0057 |
| 4 | A | .0065 |
| 5 | A | .0291 |
| 6 | O | .0174 |
| 7 | C | .0104 |
| 8 | A | .0311 |
| 9 | A | .0405 |
| 10 | E | .0064 |
| 11 | E | .0945 |
| 12 | C | .0150 |
| 13 | A | .0031 |
| 14 | A | .0097 |
| 15 | A | .0024 |
| 16 | O | .0019 |
| 17 | O | .2669* |
| 18 | E | .0158 |
| 19 | O | .0621 |
| 20 | A | .0075 |
| 21 | E | .0005 |
| 22 | C | .0119 |
| 23 | A | .0012 |
| 24 | E | .0053 |
| 25 | A | .0206 |
| 26 | A | .0031 |
| 27 | A | .0038 |
| 28 | O | .0015 |
| 29 | O | .0283 |
| 30 | N | .0900 |
| 31 | A | .0298 |
| 32 | N | .0503 |
| 33 | C | .0781 |
| 34 | C | .0057 |
| 35 | N | .0029 |
| 36 | N | .0209 |
| 37 | A | .0115 |
| 38 | E | .0425 |
| 39 | E | .0393 |

| | | |
|---|---|---|
| 40 | N | .0238 |
| 41 | E | .0207 |
| 42 | A | .0172 |
| 43 | N | .0297 |
| 44 | N | .0087 |
| 45 | C | .0006 |
| 46 | C | .0114 |
| 47 | C | .0079 |
| 48 | N | .0148 |
| 49 | N | .0036 |
| 50 | C | .0076 |
| 51 | O | .0259 |
| 52 | E | .0005 |
| 53 | N | .0026 |
| 54 | O | .0306 |
| 55 | E | .0015 |
| 56 | C | .0880 |
| 57 | C | .0019 |
| 58 | O | .0650 |
| 59 | C | .0193 |
| 60 | N | .0108 |
| 61 | N | .1503* |
| 62 | N | .0904 |
| 63 | E | .0066 |
| 64 | N | .0143 |
| 65 | O | .0014 |
| 66 | O | .0202 |
| 67 | E | .0118 |
| 68 | A | .1028* |
| 69 | O | .0012 |
| 70 | A | .0195 |
| 71 | A | .0008 |
| 72 | N | .0059 |
| 73 | A | .0449 |
| 74 | E | .0044 |
| 75 | O | .0604 |
| 76 | E | .0159 |
| 77 | O | .0089 |
| 78 | O | .0072 |
| 79 | N | .0384 |
| 80 | C | .0013 |
| 81 | E | .0149 |
| 82 | N | .0617 |
| 83 | A | .0455 |
| 84 | A | .0077 |
| 85 | O | .0390 |

| | | |
|---|---|---|
| 86 | O | .0002 |
| 87 | O | .0051 |
| 88 | C | .0011 |
| 89 | O | .0009 |
| 90 | O | .0381 |
| 91 | N | .0070 |
| 92 | E | .0083 |
| 93 | A | .0055 |
| 94 | O | .0261 |
| 95 | E | .0138 |
| 96 | C | .0156 |
| 97 | C | .0072 |
| 98 | C | .0097 |
| 99 | N | .0068 |
| 100 | N | .0196 |
| 101 | C | .0531 |
| 102 | C | .0078 |
| 103 | O | .0133 |
| 104 | O | .0324 |
| 105 | E | .0017 |
| 106 | C | .0034 |
| 107 | N | .0127 |
| 108 | C | .0175 |
| 109 | A | .0381 |
| 110 | E | .0314 |
| 111 | E | .0425 |
| 112 | E | .0025 |
| 113 | O | .0052 |
| 114 | A | .0114 |
| 115 | C | .0215 |
| 116 | C | .0088 |
| 117 | N | .0128 |
| 118 | A | .0001 |
| 119 | N | .0317 |
| 120 | C | .0524 |
| 121 | C | .0120 |
| 122 | N | .0508 |
| 123 | N | .0019 |
| 124 | E | .0232 |
| 125 | O | .0810 |
| 126 | A | .0070 |
| 127 | A | .0299 |
| 128 | C | .0047 |
| 129 | A | .0012 |
| 130 | A | .0249 |
| 131 | A | .0050 |

| | | |
|---|---|---|
| 132 | A | .0379 |
| 133 | N | .0215 |
| 134 | N | .0292 |
| 135 | N | .1471* |
| 136 | E | .0395 |
| 137 | N | .0061 |
| 138 | N | .0022 |
| 139 | C | .0004 |
| 140 | O | .0182 |
| 141 | O | .0703 |
| 142 | O | .0072 |
| 143 | A | .0142 |
| 144 | O | .0007 |
| 145 | O | .0184 |
| 146 | E | .0222 |
| 147 | A | .0760 |
| 148 | E | .0123 |
| 149 | A | .0213 |
| 150 | E | .0134 |
| 151 | C | .0023 |
| 152 | E | .0199 |
| 153 | N | .0043 |
| 154 | A | .0104 |
| 155 | E | .0023 |
| 156 | N | .0273 |
| 157 | C | .0189 |
| 158 | C | .0006 |
| 159 | C | .0124 |
| 160 | O | .0574 |
| 161 | N | .8579* |
| 162 | E | .0124 |
| 163 | C | .0258 |
| 164 | E | .0209 |
| 165 | A | .0113 |
| 166 | N | .0140 |
| 167 | C | .0128 |
| 168 | A | .0067 |
| 169 | N | .0388 |
| 170 | O | .0032 |
| 171 | E | .0000 |
| 172 | E | .0167 |
| 173 | N | .0388 |
| 174 | O | .0181 |
| 175 | O | .0502 |
| 176 | C | .0054 |
| 177 | C | .0305 |

| | | |
|---|---|---|
| 178 | E | .0063 |
| 179 | A | .0536 |
| 180 | O | .0307 |
| 181 | E | .0119 |
| 182 | N | .0022 |
| 183 | C | .0548 |
| 184 | N | .0435 |
| 185 | E | .0293 |
| 186 | C | .0004 |
| 187 | E | .0328 |
| 188 | N | .0455 |
| 189 | A | .0332 |
| 190 | E | .0481 |
| 191 | A | .0149 |
| 192 | N | .0022 |
| 193 | E | .0077 |
| 194 | O | .0017 |
| 195 | O | .0156 |
| 196 | E | .0169 |
| 197 | E | .0448 |
| 198 | E | .0447 |
| 199 | A | .0011 |
| 200 | E | .0299 |
| 201 | C | .0066 |
| 202 | O | .0081 |
| 203 | N | .0010 |
| 204 | C | .0017 |
| 205 | E | .0005 |
| 206 | N | .0551 |
| 207 | C | .0060 |
| 208 | O | .0464 |
| 209 | E | .0134 |
| 210 | C | .0151 |
| 211 | O | .0130 |
| 212 | E | .0156 |
| 213 | C | .0026 |
| 214 | C | .0112 |
| 215 | O | .0007 |
| 216 | N | .0124 |
| 217 | N | .0073 |
| 218 | N | .0068 |
| 219 | C | .0194 |
| 220 | E | .0303 |
| 221 | O | .0169 |
| 222 | N | .0121 |
| 223 | O | .0162 |

| | | |
|---|---|---|
| 224 | O | .0095 |
| 225 | A | .0003 |
| 226 | N | .0003 |
| 227 | A | .0036 |
| 228 | O | .0207 |
| 229 | E | .0022 |
| 230 | E | .0172 |
| 231 | C | .0113 |
| 232 | A | .0310 |
| 233 | N | .0126 |
| 234 | C | .0136 |
| 235 | O | .0287 |
| 236 | A | .1311* |
| 237 | E | .0002 |
| 238 | A | .0510 |
| 239 | O | .0394 |
| 240 | A | .0105 |
| 241 | N | .0227 |
| 242 | A | .0087 |
| 243 | N | .0144 |
| 244 | E | .0052 |
| 245 | C | .0009 |
| 246 | E | .0030 |
| 247 | N | .0227 |
| 248 | C | .0081 |
| 249 | A | .0128 |
| 250 | A | .0014 |
| 251 | A | .0479 |
| 252 | O | .0373 |
| 253 | C | .0043 |
| 254 | N | .0062 |
| 255 | N | .0019 |
| 256 | C | .0092 |
| 257 | E | .0207 |
| 258 | O | .0008 |
| 259 | A | .0751 |
| 260 | O | .0077 |
| 261 | A | .0018 |
| 262 | A | .0674 |
| 263 | N | .0048 |
| 264 | O | .0040 |
| 265 | N | .0799 |
| 266 | E | .0118 |
| 267 | C | .0025 |
| 268 | O | .0437 |
| 269 | A | .0029 |

| | | |
|---|---|---|
| 270 | E | .0168 |
| 271 | C | .0339 |
| 272 | E | .0433 |
| 273 | E | .0259 |
| 274 | C | .0075 |
| 275 | O | .0006 |
| 276 | A | .0057 |
| 277 | O | .0103 |
| 278 | N | .0521 |
| 279 | A | .0864 |
| 280 | A | .0204 |
| 281 | N | .0099 |
| 282 | O | .0028 |
| 283 | O | .0054 |
| 284 | A | .0005 |
| 285 | O | .0415 |
| 286 | C | .0214 |
| 287 | N | .0045 |
| 288 | E | .0365 |
| 289 | E | .0076 |
| 290 | C | .0477 |
| 291 | O | .0029 |
| 292 | N | .0090 |
| 293 | O | .1908* |
| 294 | C | .0099 |
| 295 | C | .0031 |
| 296 | C | .0031 |
| 297 | A | .0056 |
| 298 | E | .0112 |
| 299 | C | .0168 |
| 300 | E | .0086 |

Note. C=Conscientiousness, E=Extraversion, N= Neuroticism, O=Openness, A= Agreeableness.

Items donated by * are identified as functioning differentially.

Table 13

*Manipulation Check of Item Type Condition Comparing Men and Women Scores*

| Stereotype | Item | Men | | Women | | |
|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | *t*-test |
| Feminine | Get overwhelmed by emotions. | 2.91 | 1.06 | 3.53 | 1.00 | -4.25** |
| Neutral | Get impacted by moods. | 3.44 | 1.07 | 3.68 | 0.89 | -1.75 |
| Masculine | Not overwhelmed by emotions. | 3.07 | 1.02 | 2.83 | 1.14 | 1.61 |
| Feminine | Experience my emotions intensely. | 3.21 | 1.01 | 3.59 | 1.08 | -2.64* |
| Neutral | Experience my emotions. | 3.63 | 0.90 | 3.85 | 0.96 | -1.71 |
| Masculine | Don't notice my emotions. | 2.50 | 1.03 | 2.38 | 1.18 | 0.84 |
| Masculine | Am not easily affected by my emotions. | 2.91 | 1.12 | 2.73 | 1.15 | 1.10 |
| Neutral | Sometimes act based on my emotions. | 3.53 | 0.98 | 3.60 | 1.01 | -0.55 |
| Feminine | Am easily affected by my emotions. | 3.03 | 1.12 | 3.45 | 1.12 | -2.68* |
| Masculine | Don't understand people who get emotional. | 2.47 | 1.05 | 2.20 | 1.09 | 1.81 |
| Neutral | Don't understand people who are affected by their feelings. | 2.46 | 0.98 | 2.22 | 1.09 | 1.68 |
| Feminine | Sympathize with those who get emotional. | 3.59 | 0.92 | 3.79 | 1.11 | -1.41 |
| Communal | Value cooperation over competition. | 3.71 | 0.93 | 3.74 | 0.92 | -0.22 |
| Neutral | Value cooperation and competition. | 3.86 | 0.87 | 3.74 | 0.99 | 0.93 |
| Agentic | Value competition and cooperation. | 3.88 | 1.01 | 3.66 | 0.95 | 1.63 |
| Agentic | Remain calm under pressure. | 3.60 | 1.01 | 3.47 | 1.02 | 0.96 |
| Neutral | Remain at ease under pressure. | 3.40 | 1.03 | 3.23 | 1.07 | 1.17 |
| Communal | Help others stay calm under pressure. | 3.66 | 0.86 | 3.76 | 0.96 | -0.81 |
| Agentic | Get chores done right away. | 2.86 | 1.12 | 3.06 | 1.04 | -1.34 |
| Neutral | Get chores done. | 3.53 | 0.97 | 3.70 | 0.86 | -1.37 |
| Communal | Get mine and others chores done right away. | 2.84 | 1.11 | 2.96 | 1.05 | -0.80 |
| Agentic | Think highly of myself. | 3.48 | 1.05 | 3.19 | 1.09 | 1.90* |
| Neutral | Think well of myself. | 3.65 | 0.94 | 3.48 | 1.09 | 1.16 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Communal | Think highly of how I care for others. | 3.67 | 0.96 | 3.86 | 0.90 | -1.44 |
| Warm | Am relaxed most of the time. | 3.64 | 0.96 | 3.09 | 1.07 | 3.94** |
| Neutral | Am sometimes relaxed. | 3.74 | 0.99 | 3.62 | 0.94 | 0.89 |
| Not Warm | Am often anxious. | 2.96 | 1.08 | 3.46 | 1.11 | -3.32** |
| Warm | Rarely complain. | 3.11 | 1.13 | 2.67 | 1.08 | 2.89** |
| Neutral | Sometimes I complain. | 3.41 | 1.06 | 3.84 | 0.85 | -3.25** |
| Not Warm | Often complain. | 2.71 | 1.15 | 3.03 | 1.18 | -1.96* |
| Competent | Like order. | 3.63 | 0.87 | 3.94 | 0.88 | -2.51* |
| Neutral | Don't mind order. | 3.54 | 0.88 | 3.54 | 1.02 | 0.04 |
| Not Competent | Dislike order. | 2.59 | 1.07 | 2.30 | 1.04 | 1.99* |
| Competent | Easily resist temptations. | 2.97 | 0.99 | 3.30 | 0.95 | -2.48* |
| Neutral | Somewhat easily resist temptations. | 3.14 | 0.87 | 3.32 | 0.89 | -1.51 |
| Not Competent | Unable to resist temptations. | 2.81 | 1.05 | 2.72 | 1.04 | 0.62 |
| Competent | Am calm in even tense situations. | 3.43 | 1.02 | 3.46 | 1.04 | -0.22 |
| Neutral | Somewhat calm in even tense situations. | 3.59 | 0.87 | 3.43 | 0.99 | 1.25 |
| Not Competent | Am not calm in tense situations. | 2.68 | 1.01 | 2.66 | 1.12 | 0.17 |
| Competent | Adapt easily to new situations. | 3.43 | 0.89 | 3.46 | 1.06 | 0.49 |
| Neutral | Adapt to new situations. | 3.81 | 0.77 | 3.75 | 1.02 | 0.45 |
| Not Competent | Can't successfully adapt to new situations. | 2.38 | 0.98 | 2.43 | 1.17 | -0.34 |
| Competent | Know the answers to many questions. | 3.09 | 0.90 | 3.26 | 0.92 | -1.29 |
| Neutral | Am able to answer some questions. | 3.91 | 0.76 | 4.00 | 0.71 | -0.93 |
| Not Competent | Often don't know the answers to many questions. | 2.83 | 1.02 | 2.98 | 1.09 | -1.03 |
| Competent | Excel in what I do. | 3.60 | 0.88 | 3.60 | 0.88 | 0.05 |
| Neutral | Excel in some things that I do. | 3.79 | 0.84 | 3.80 | 0.83 | -0.04 |
| Not Competent | Don't excel at anything I do. | 2.23 | 1.00 | 2.24 | 1.04 | -0.07 |
| Warm | Radiate joy. | 3.25 | 1.03 | 3.42 | 1.03 | -1.19 |
| Neutral | Occasionally joyous. | 3.70 | 0.81 | 3.63 | 1.08 | 0.47 |
| Not Warm | Don't often feel joy. | 2.50 | 1.09 | 2.51 | 1.15 | -0.04 |
| Warm | Trust what people say. | 3.22 | 0.95 | 3.33 | 0.99 | -0.87 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Neutral | Sometimes trust what people say. | 3.58 | 0.84 | 3.59 | 0.90 | -0.12 |
| Not Warm | Never trust what people say. | 2.69 | 1.04 | 2.51 | 1.10 | 1.24 |
| Competent | Do more than what's expected of me. | 3.28 | 0.98 | 3.65 | 0.91 | -2.79* |
| Neutral | Do what is generally expected of me. | 3.79 | 0.83 | 3.87 | 0.85 | -0.68 |
| Not Competent | Do less than what is expected of me. | 2.56 | 0.98 | 2.19 | 0.97 | 2.70* |
| Warm | Amuse my friends. | 3.92 | 0.97 | 3.76 | 0.85 | 1.24 |
| Neutral | Entertain others. | 3.76 | 0.94 | 3.60 | 0.92 | 1.19 |
| Not Warm | Don't amuse my friends. | 2.21 | 1.03 | 2.10 | 0.87 | 0.82 |
| Warm | Think that all will be well. | 3.60 | 0.92 | 3.58 | 1.04 | 0.17 |
| Neutral | Often think that all will end up ok. | 3.67 | 0.99 | 3.48 | 1.02 | 1.35 |
| Not Warm | Hardly ever think that all will be well. | 2.56 | 1.04 | 2.56 | 1.13 | 0.01 |
| Warm | Involve others in what I am doing. | 3.41 | 0.96 | 3.54 | 0.97 | -0.99 |
| Neutral | Sometimes involve others in what I am doing. | 3.66 | 0.81 | 3.49 | 0.99 | 1.30 |
| Not Warm | Prefer to do things alone. | 3.45 | 1.08 | 3.46 | 1.07 | -0.09 |

Note *(*p<.05, **p<.001)*

Table 14

*Results of Mixed ANOVA by Item Type*

| Personality Factor | Stereotype | Item Pair (Original/Opposite) | Main Effect | | Interaction | | Interaction of Gender |
|---|---|---|---|---|---|---|---|
| | | | $\eta_\rho 2$ | F | $\eta_\rho 2$ | F | |
| Conscientiousness | Agentic/Communal | Get chores done right away. /Get mine and others chores done right away. | 0.005 | 0.99 | 0.002 | 0.37 | N |
| Conscientiousness | Competent/ Not Competent | Like order. /Dislike order. | 0.418 | 148.81** | 0.035 | 7.52* | Y |
| Conscientiousness | Competent/Not Competent | Excel in what I do./Don't excel at anything I do. | 0.431 | 155.43** | 0.000 | 0.00 | N |
| Conscientiousness | Competent/ Not Competent | Do more than what's expected of me. /Do less than what is expected of me. | 0.333 | 103.54** | 0.055 | 11.94** | Y |
| Conscientiousness | Feminine/Masculine | Get overwhelmed by emotions./Not overwhelmed by emotions. | 0.023 | 4.81* | 0.054 | 11.86** | Y |
| Extraversion | Warm/Not Warm | Radiate joy. /Don't often feel joy. | 0.178 | 44.31** | 0.002 | 0.49 | N |
| Extraversion | Warm/Not Warm | Amuse my friends. /Don't amuse my friends. | 0.529 | 231.81** | 0.000 | 0.08 | N |
| Extraversion | Warm/Not Warm | Involve others in what I am doing. /Prefer to do things alone. | 0.000 | 0.02 | 0.002 | 0.32 | N |
| Openness | Feminine/Masculine | Experience my emotions intensely/Don't notice my emotions. | 0.263 | 73.67** | 0.025 | 5.21* | Y |
| Openness | Masculine/Feminine | Am not easily affected by my emotions. /Am easily affected by my emotions. | 0.050 | 10.86** | 0.025 | 5.26* | Y |
| Openness | Masculine/Feminine | Don't understand people who get emotional. /Sympathize with those who get emotional. | 0.378 | 125.33** | 0.019 | 4.01* | Y |
| Agreeableness | Communal/Agentic | Value cooperation over competition. /Value competition and cooperation. | 0.001 | 0.22 | 0.011 | 2.29 | N |
| Agreeableness | Agentic/Communal | Think highly of myself. /Think highly of how I care for others. | 0.104 | 24.03** | 0.037 | 7.81* | Y |
| Agreeableness | Competent/Not Competent | Know the answers to many questions. /Often don't know the answers to many questions. | 0.029 | 6.24* | 0.000 | 0.00 | N |
| Agreeableness | Warm/Not Warm | Trust what people say. /Never trust what people say. | 0.134 | 31.99** | 0.008 | 1.59 | N |
| Agreeableness | Warm/Not Warm | Think that all will be well. /Hardly ever think that all will be well. | 0.259 | 72.21** | 0.000 | 0.08 | N |
| Neuroticism | Competent/Not Competent | Easily resist temptations. /Unable to resist temptations. | 0.048 | 10.48** | 0.017 | 3.51 | N |
| Neuroticism | Competent/ Not Competent | Am calm in even tense situations. /Am not calm in tense situations. | 0.167 | 41.16** | 0.000 | 0.05 | N |
| Neuroticism | Competent/ Not Competent | Adapt easily to new situations. /Can't successfully adapt to new situations. | 0.313 | 93.55** | 0.001 | 0.29 | N |
| Neuroticism | Warm/Not Warm | Am relaxed most of the time. /Am often anxious. | 0.009 | 1.79 | 0.094 | 21.60** | Y |
| Neuroticism | Warm/Not Warm | Rarely complain. /Often complain. | 0.000 | 0.01 | 0.036 | 7.79* | Y |
| Neuroticism | Agentic/Communal | Remain calm under pressure. / Help others stay calm under pressure. | 0.030 | 6.236** | 0.012 | 0.37 | N |

Note *(*p<.05, **p<.001)*

Figure 1
*Option Characteristic Curve*

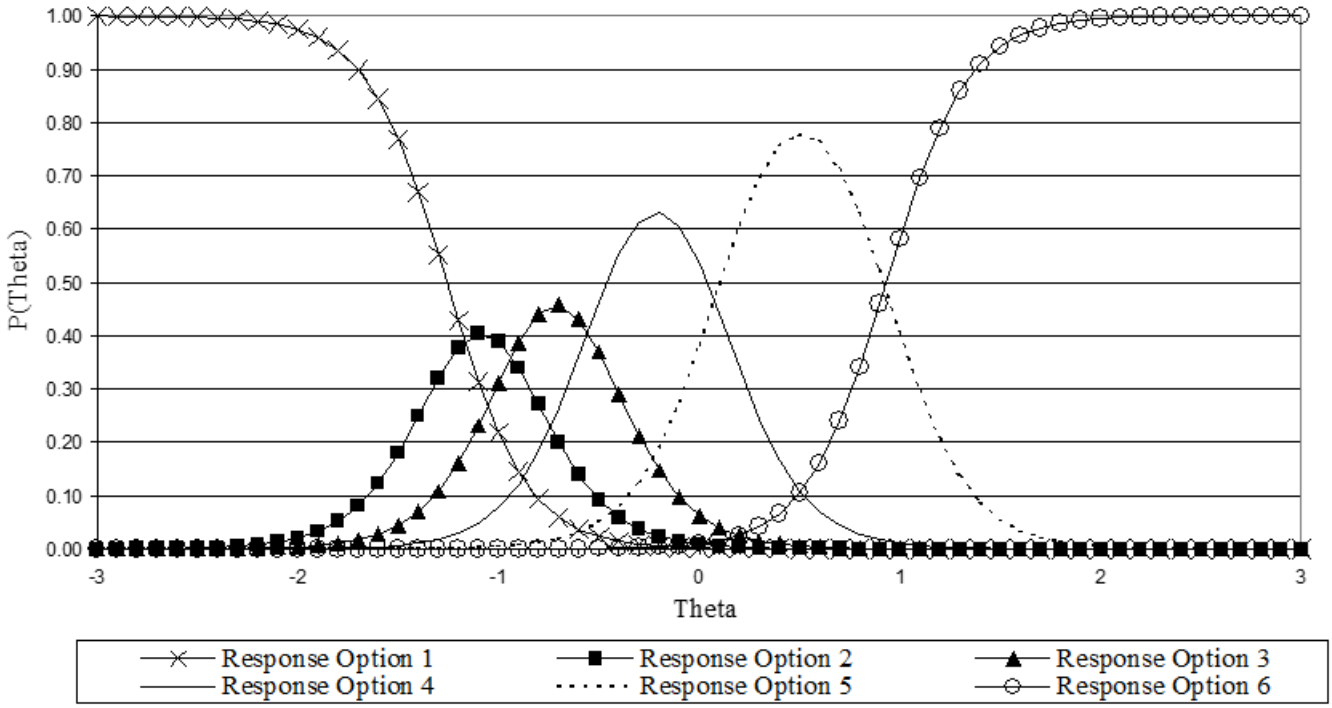


Figure 2

*Sample Interaction Plot (Study 2)*

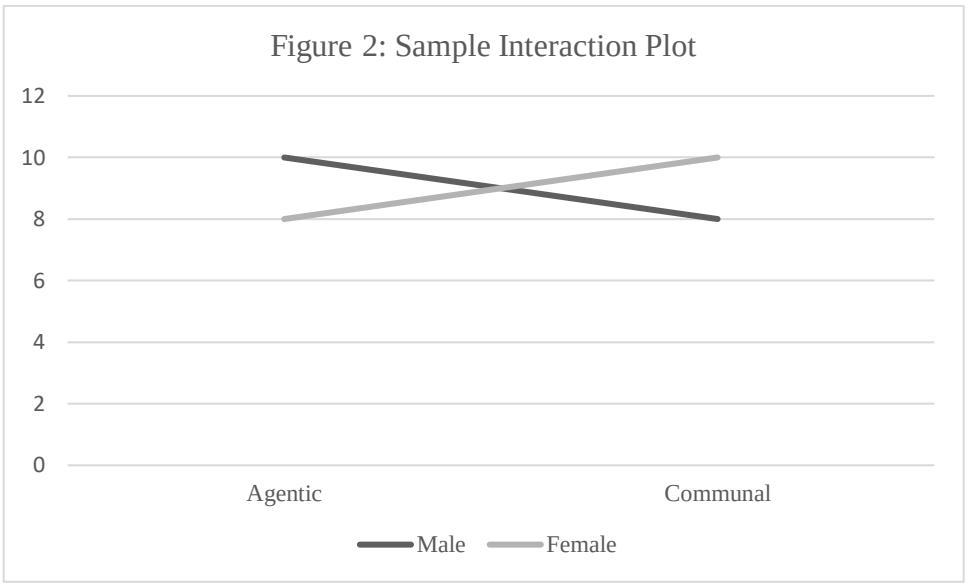


Figure 3

*Plot of Neuroticism Item 61 with DIF under the GRM model parameter between groups*

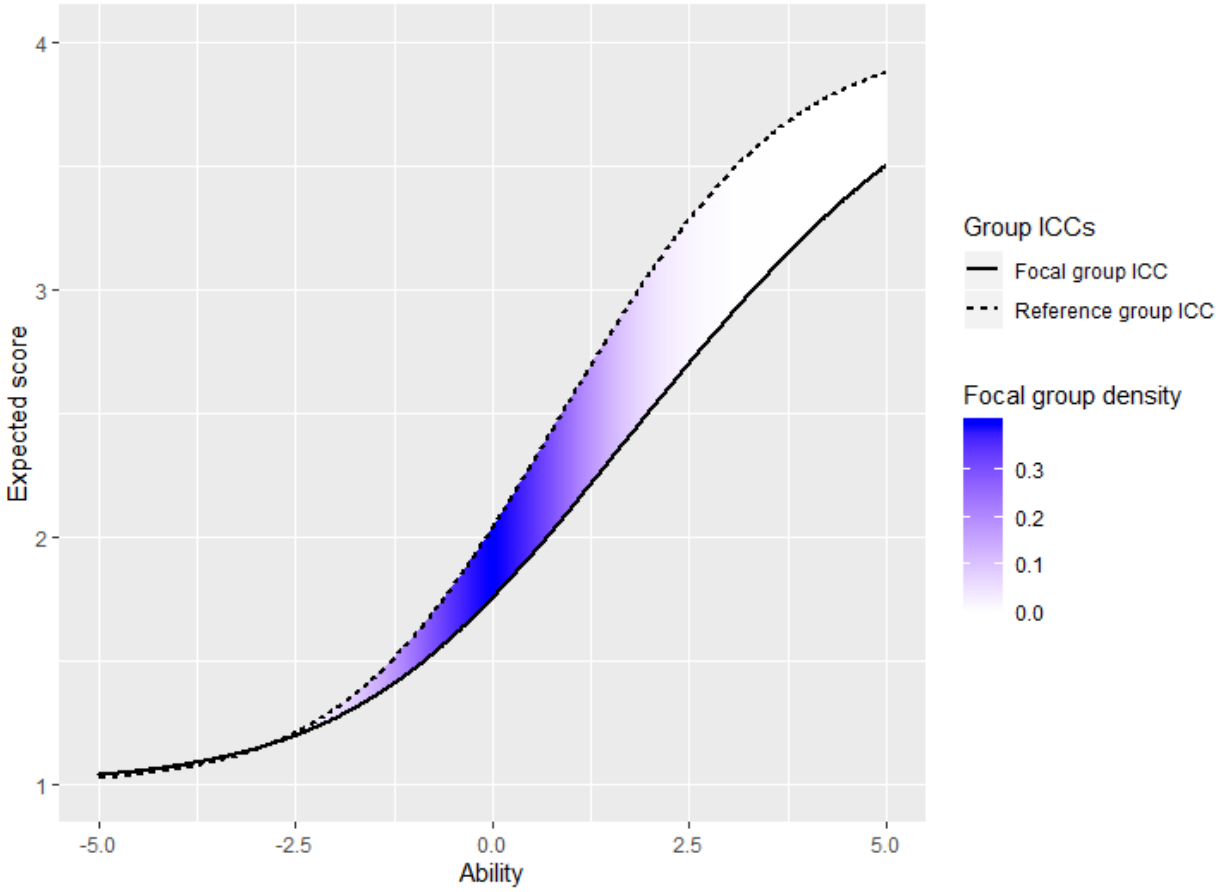


Figure 4

*Plot of Neuroticism Item 135 with DIF under the GRM model parameter between groups*

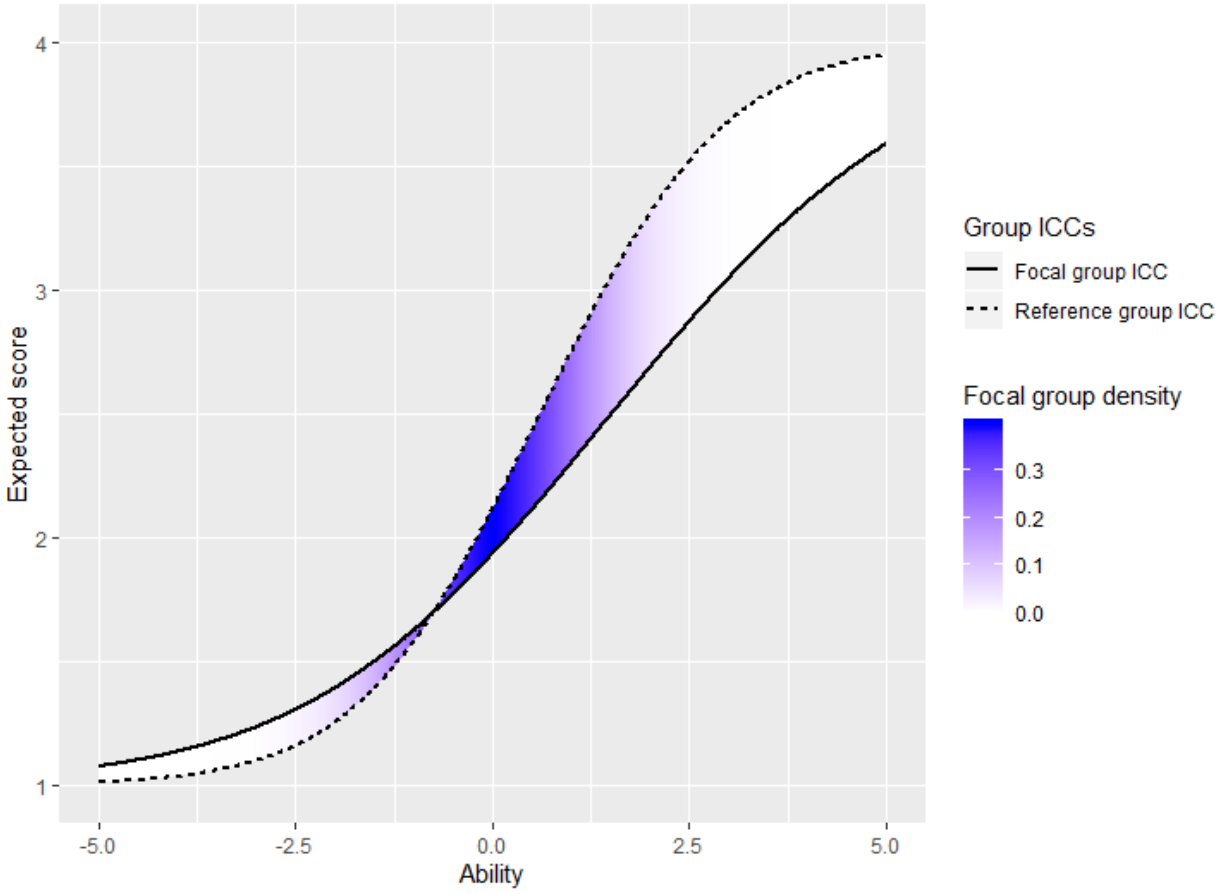


Figure 5

*Plot of Neuroticism Item 161 with DIF under the GRM model parameter between groups*

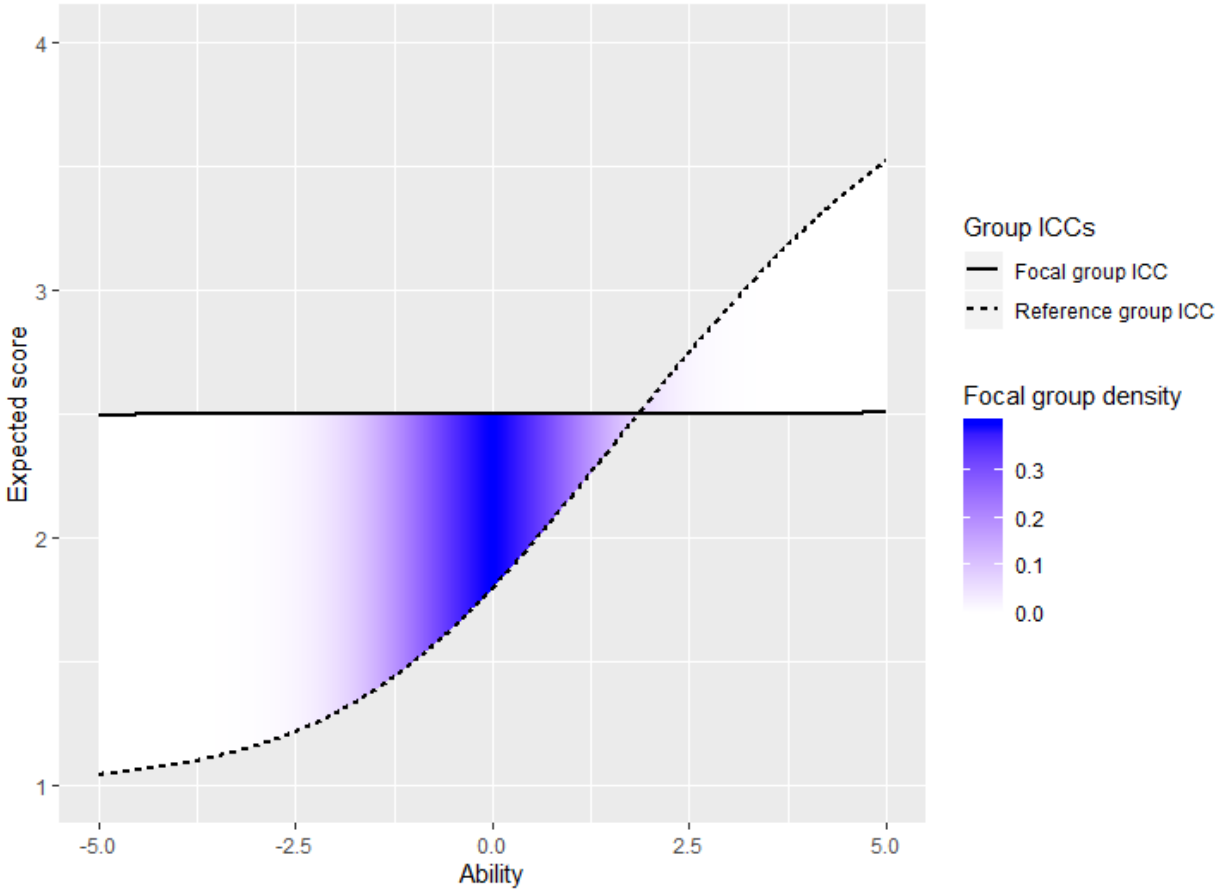


Figure 6

*Plot of Agreeableness Item 68 with DIF under the GRM model parameter between groups*

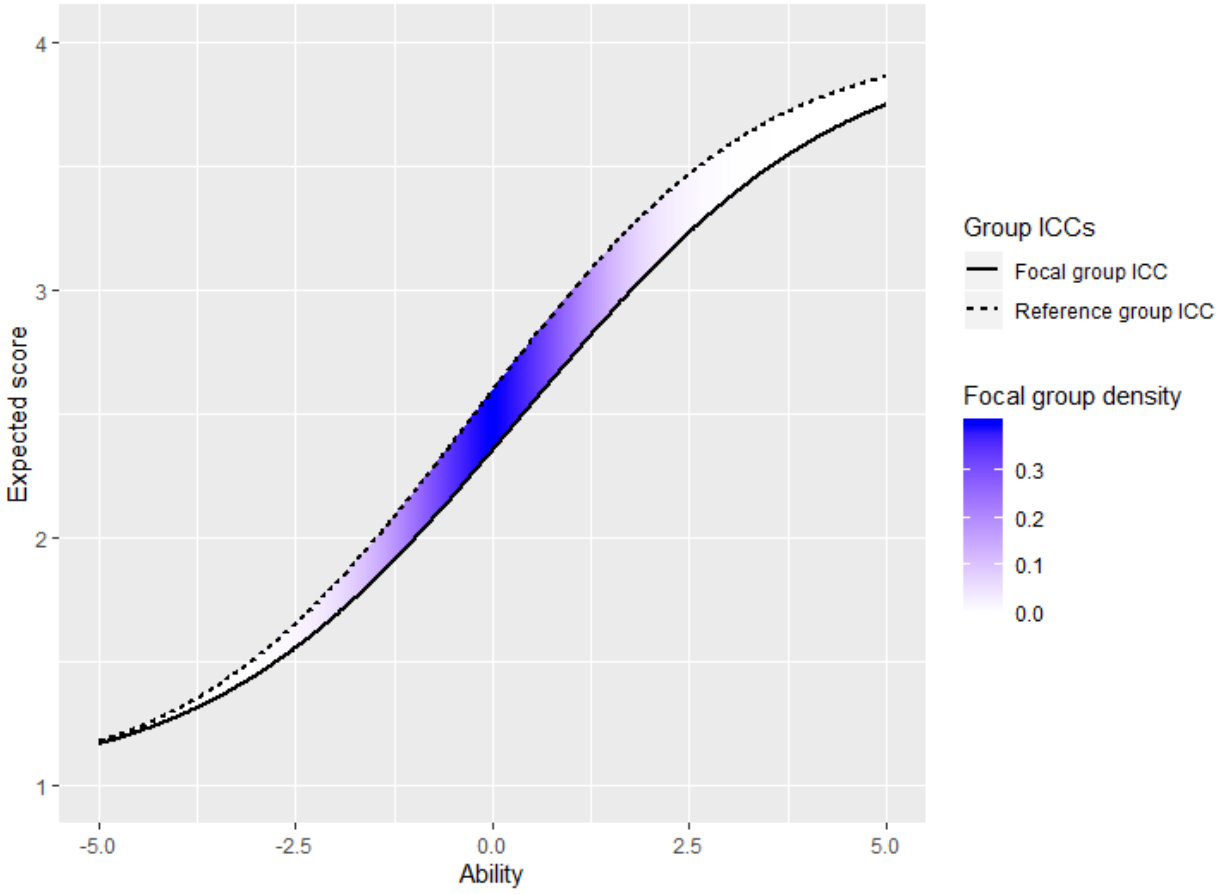


Figure 7

*Plot of Agreeableness Item 236 with DIF under the GRM model parameter between groups*

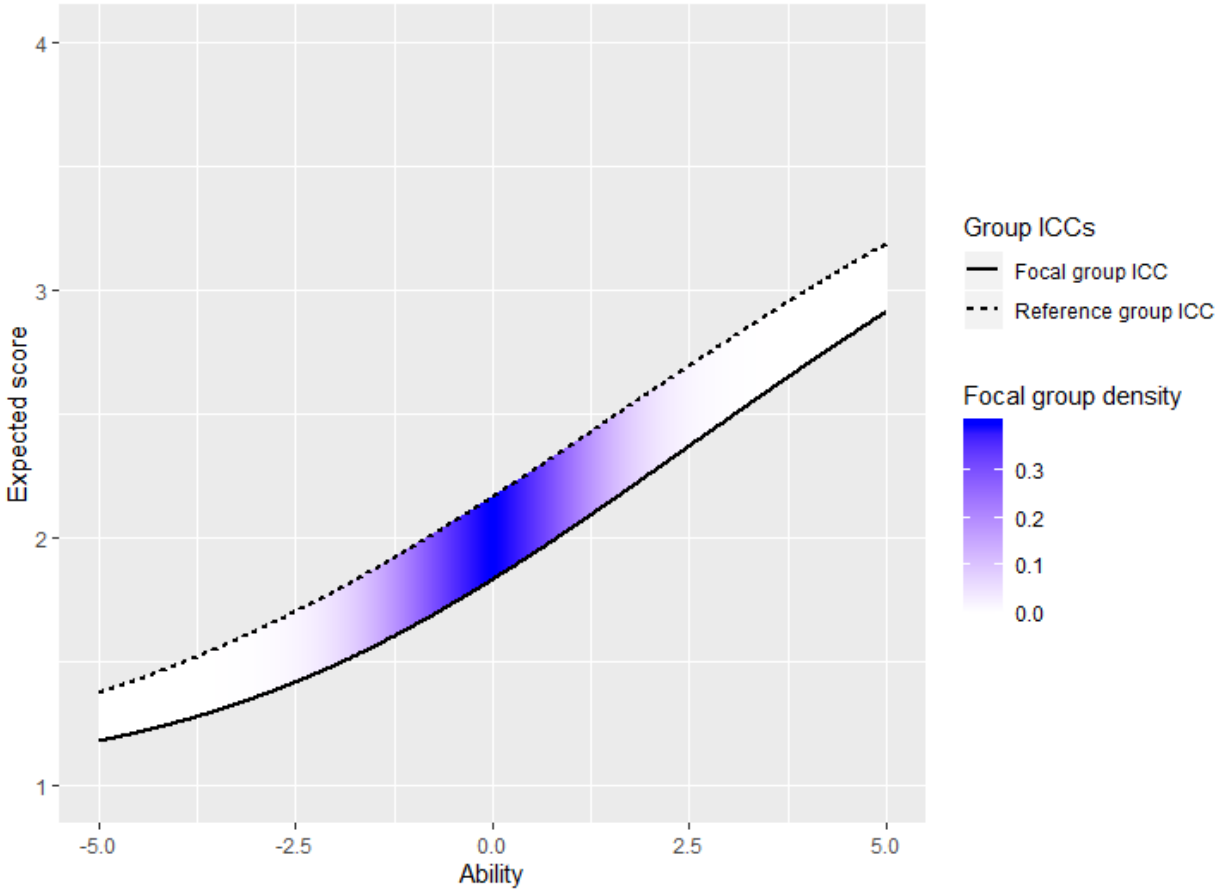


Figure 8

*Plot of Openness Item 17 with DIF under the GRM model parameter between groups*

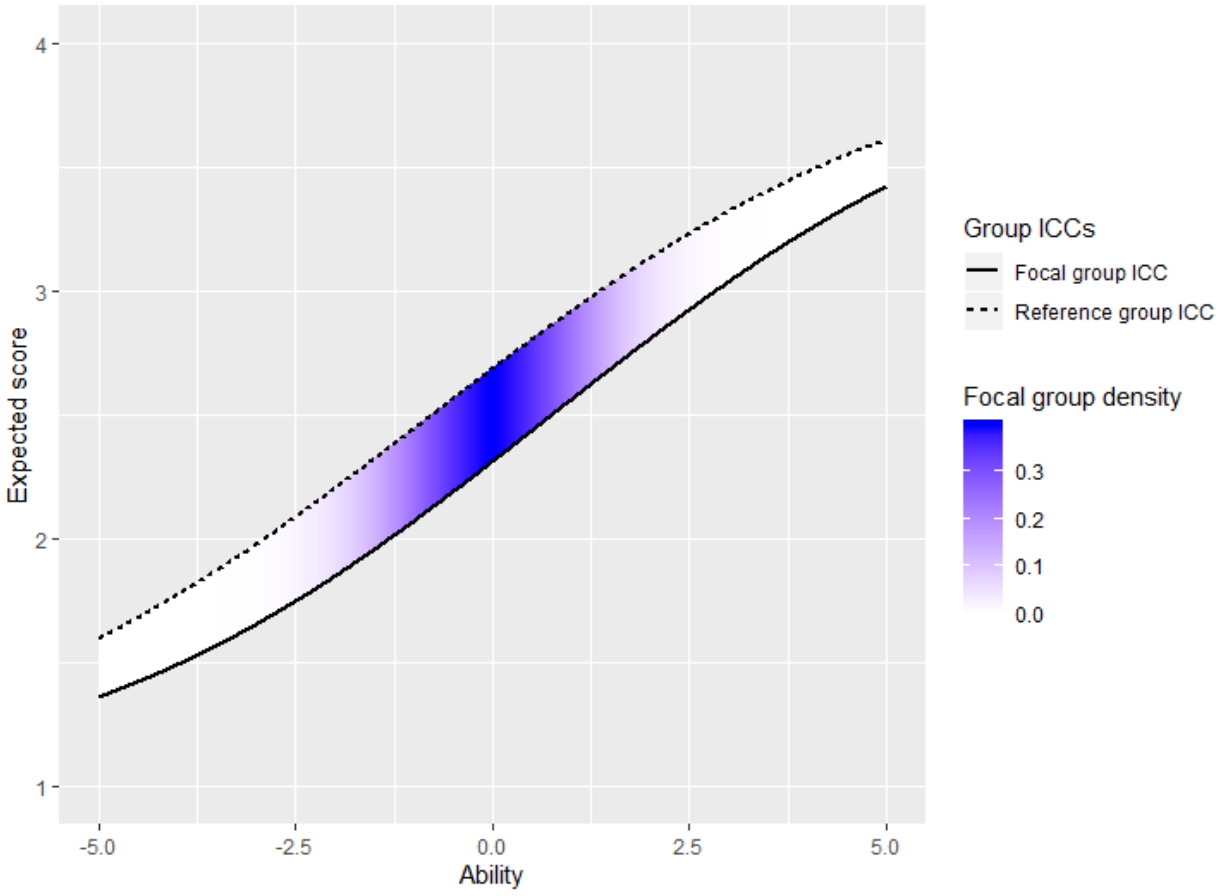
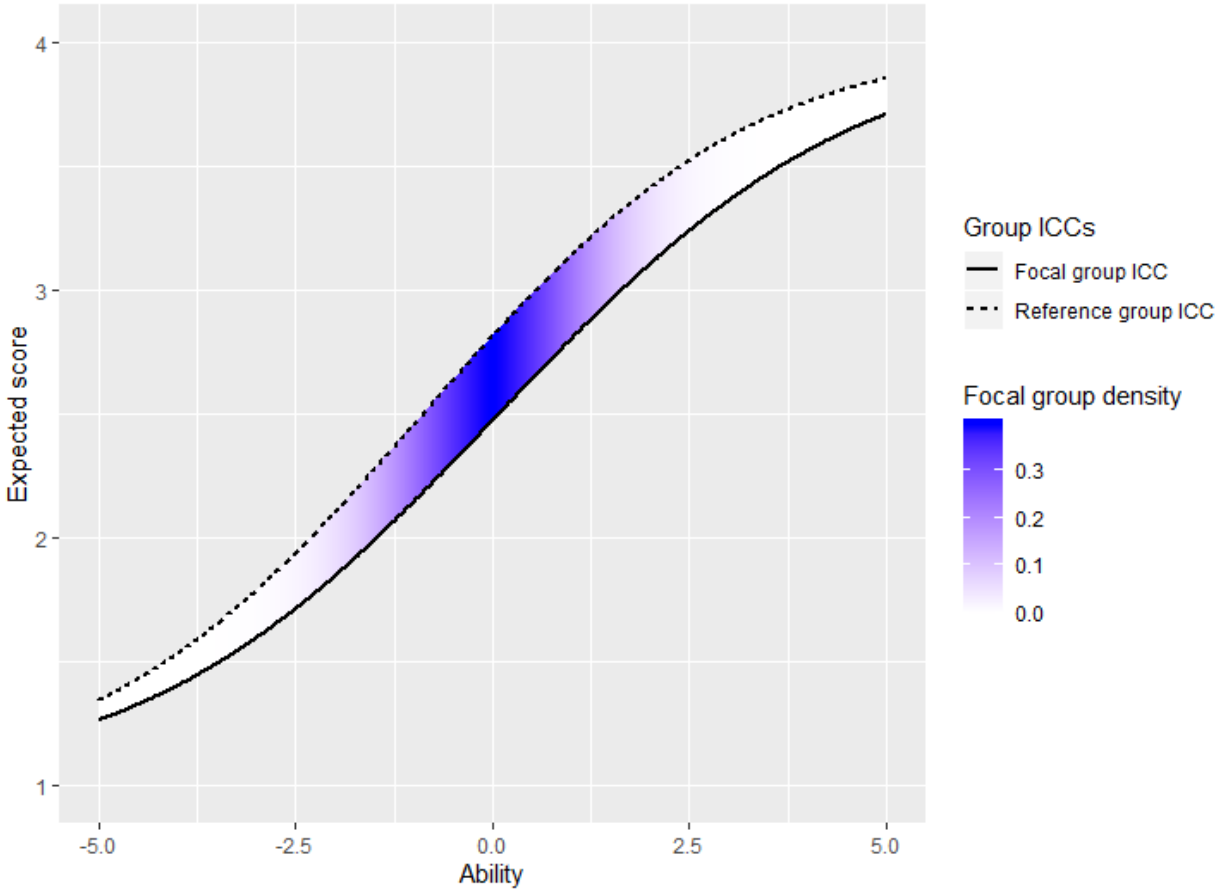


Figure 9

*Plot of Openness Item 60 with DIF under the GRM model parameter between groups*



**References**

Allport, G. W. (1921). Personality and character. *Psychological Bulletin, 18,* 441-455.

Allport, G.W. (1937). Personality: A psychological interpretation. Oxford, England: Holt.

Allport, G.W. (1954). *The nature of prejudice.* Reading, MA: Addison-Wesley.

Allport, G.W. (1996). Traits revisited. *American Psychologist, 21,* 1-10.

Allport, G.W. & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47,* i-171.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist, 38,* 1145-1160.

Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man*. Chicago: Rand McNally.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Barrick, M.R., & Mount, M.K. (1991). The Big Five personality dimensions and job Performance: A meta-analysis. *Personnel Psychology, 44,* 1-26.

Barrick, M.R., Mount, M.K., & Judge, T.A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance, 9,* 9-30.

Baumeister, R. F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspectives on Psychological Science, 2,* 396-403.

Bem, S.L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155-162.

Benet-Martinez, V., Donnellan, M. B., Fleeson, W., Fraley, C., Gosling, S.D., King, L.A., Robins, R.W., Funder, D. C. (2014). Six visions for the future of personality psychology. In M. Mikulincer, P.R., Shaver, M. L., Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Vol 4. Personality processes and individual differences (pp. 665-689).* Washington, DC, US: American Psychological Association.

Bianchi, S. M., Sayer, L. C., Milkie, M. A., & Robinson, J. P. (2012). Housework: Who did, does or will do it, and how much does it matter? *Social Forces: A Scientific Medium of Social Study and Interpretation, 91,* 55-63.

Block, J. H. (1973). Conceptions of sex role: Some cross-cultural and longitudinal perspectives. *American Psychologist, 28,* 512-526.

Boyle, G, J. & Helmes, E. (2009). Methods of personality assessment. In P.J. Corr & G. Matthews (Eds.), *The Cambridge Handbook of Personality Psychology* (pp. 110-126). Cambridge: Cambridge University Press.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P.S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues, 28,* 59-78.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry, 6,* 1-30.

Buss, D. M. (2007). The evolution of human mating. *Acta Psychologica Sinica, 39,* 502-512.

Buss, D. M. (2009). How can evolutionary psychology successfully explain personality and individual differences? *Perspectives on Psychological Science, 4*, 359-366.

Buss, D. & Penke, L. (2015). Evolutionary personality psychology. In M. Mikulincer, P.R., Shaver, M. L., Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Vol 4. Personality processes and individual differences (pp. 3-29).*

Washington, DC, US: American Psychological Association.

Carter, M. J. (2014). Gender socialization and identity theory. *Social Sciences, 3,* 242-263.

Cattell, R. B. (1946). Description and measurement of personality. Oxford: World Book Company.

Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika, 12,* 197-220.

Cervantes, V. H. (2017). DFIT: An R package for Raju's differential functioning of items and tests framework. *Journal of Statistical Software, 76,* 1-24.

Christiansen, N. D. & Tett, R. P. (2013). *Handbook of Personality at Work.* London: Routledge.

Cohen, A. S., & Kim, S. (1998). An investigation of linking methods under the graded response model. *Applied Psychological Measurement, 22,* 116-130.

Costa, P.T., Jr. (1996). Work and personality: Use of the NEO-PI-R in Industrial/ Organizational psychology. *Applied Psychology: An International Review, 45,* 225-241.

Costa, P. T. Jr & McCrae, R. R. (1997). Stability and change in personality assessment: The Revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68,* 86-94.

Costa, P.T., Jr. & McCrae, R. R. (1998). Six approaches to the explication of facet-level traits: Examples from conscientiousness. *European Journal of Personality, 12,* 117-134.

Costa, P. T., Jr. & McCrae, R.R. (2008). The revised NEO Personality Inventory (NEO-PI-R). In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *Sage Handbook of Personality Theory and Assessment, Volume 2,* (pp. 179-198). Los Angeles: Sage.

Costa, P. T., Jr., & McCrae, R. R. (2009). The Five-Factor Model and the NEO inventories. In J.N. Butcher (Ed.), *Oxford Handbook of Personality Assessment* (pp. 299-322). New

York, NY: Oxford University Press.

Costa, P. T., Jr., 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.

Church, A.H., Fleck, C.R., Foster, G.C., Levine, R.C., Lopez, F.J., &Rotolo, C. T. (2016). Does purpose matter? The stability of personality assessments in organization development and talent management applications over time. *Journal of Applied Behavioral Science, 52,* 450-481.

Costa, P. T. & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4,* 5-13.

Cronbach, L.J & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS Map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92,* 631-648.

Cuddy, A. J.C, Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The Stereotype Content Model and the BIAS Map. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (pp. 61-149). New York, NY: Academic Press.

Cuddy, A.J.C., Fiske, S.T., Kwan, V.S.Y., Glick, P., Demoulin, S., Leyens, J. P., & Ziegler, R. (2009). Is the stereotype content model culture-bound? A cross-cultural comparison reveals systematic similarities and differences. *British Journal of Social Psychology, 48,* 1-33.

Donnellan, M. B., Hill, P. L., & Roberts, B. W. (2015). Personality development across the life

span: Current findings and future directions. In M. Mikulincer and P. R.Shaver (Eds.), APA *Handbook of Personality and Social Psychology: Personality Processes and Individual Differences, 4*, 107-126.

Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.

Drasgow, F. & Hulin, C. L. (1990). Item response theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 577-636). Palo Alto, CA: Consulting Psychologists Press.

Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice test. *Applied Psychological Measurement, 19*, 143-165.

Dudley, N.M., Orvis, K.A., Lebiecki, J.E., & Cortina, J.M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits, *Journal of Applied Psychology, 91,* 40-57.

Eagly, A.H. (1987). *Sex differences in social behavior: A social-role interpretation.* Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H. (1992). Gender and helping behavior: A meta-analytic review of the social psychological literature. In C. N. Jacklin (Ed.), *Reprinted in the psychology of gender (Part of International Library of Critical Writings in Psychology* (pp. 3-28). Aldershot, England: Edward Elger.

Eagly, A.H. & Johannesen-Schmidt, M.C. (2001). The leadership styles of women and men. *Journal of Social Issues, 57*, 781-797.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2013. *American Psychologist, 75,* 301-315.

Eagly, A. H. & Sczesny, S. (2009). Stereotypes about women, men, and leaders: Have times changed? In M. Barreto, M. K. Ryan, & M. T. Schmitt (Eds.), *Psychology of women book series. The glass ceiling in the 21st century: Understanding barriers to gender equality* (pp. 21-47). Washington, DC: American Psychological Association.

Eagly, A. H. & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46,* 735-754.

Eagly, A.H. & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin, 17,* 306-315.

Eagly, A.H. & Wood, W. (2001). The origins of sex differences in human behavior: Evolved dispositions versus social roles. In D. Funder, & D.J. Ozer (Eds.), *Reprinted in Pieces of the personality puzzle* (3rd ed., pp 192-210). New York: W.W. Norton.

Eagly, A.H. & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins, *Handbook of Theories of Social Psychology, Volume 2 (pp. 458-476).* Los Angeles: Sage Publications Inc.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123-174). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Eccles, J. S. (1993). School and family effects on the ontogeny of children's interests, self-perceptions, and activity choice. In *Nebraska symposium on motivation* (pp. 145-208).

Lincoln, NE: U. Nebraska Press.

Eccles, J.S. (2001). Gender and achievement. In J. Worell (Ed.), *Encyclopedia of women and gender: Sex similarities and differences and the impact of society on gender* (pp. 43-53). San Diego, CA: Academic Press.

Eccles, J.S., Jacobs, J.E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46,* 183-201.

Eckes, T. (1994). Features of men, features of women: Assessing stereotypic beliefs about gender subtypes. *British Journal of Social Psychology, 33,* 107-123.

Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles, 47,* 99-114.

Eckes, T. & Trautner, H. M. (2000). Developmental social psychology of gender: An integrative framework. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 3-32). Mahwah, NJ: Erlbaum.

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.

Eysenck, H.J. (1991). Dimensions of personality: 16, 5, or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences, 12,* 773-790.

Fausto-Sterling, A. (1985). The new research on women: How does it affect the natural sciences? *Women's Studies Quarterly, 13,* 30-32.

Feingold, A. (1992). Gender differences in mate selection preferences: A test of the parental investment model. *Psychological Bulletin, 112,* 125-139.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin,*

*116,* 429-456.

Fishman, J. A. (1956). An examination of the process and function of social stereotyping. *Journal of Social Psychology, 43,* 27-64.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2002). Emotions up and down: Intergroup emotions result from perceived status and competition. In D. M. Mackie & E. R. Smith (Eds.), *From prejudice to intergroup emotions: Differentiated reactions to social groups* (pp. 247-264). New York: Psychology Press.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social perception: Warmth and competence. *Trends in Cognitive Science, 11,* 77-83.

Fiske, S. T., Cuddy, A.J.C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878-902.

Fiske, S.T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues, 55,* 473-489.

Fiske, S. T. (2012). The Continuum Model and the Stereotype Content Model. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of Social Psychology: Volume 1 (pp. 267-288).* London: SAGE Publications Ltd.

Fleeson, W. & Jayawickreme, E. (2014). Whole trait theory. *Journal of Research in Personality, 56,* 82-92.

Fleishman, E. A. & Quaintance, M. K. (1984). *Taxonomies of human performance*. Orlando: Academic Press.

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.

Freud, S. (1964). *An outline of psychonalysis, the standard edition of the complete psychological works of Sigmund Freud* (Vol. XXIII).

Furr, R. M. & Bacharach, V. R. (2013). *Psychometrics: An introduction* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Gaucher, D., Friesen, J. P., Neufeld, K. H. S., & Esses, V. M. (2017). Changes in the positivity of migrant stereotype content: How system-sanctioned pro-migrant ideology can affect public opinions of migrants. *Social Psychological and Personality Science, 9,* 223-233.

Ghiselli, E. (1966). *The validity of occupational aptitude tests.* New York: Wiley.

Glick, P., Diebold, J., Bailey-Werner, B., & Zhu, L. (1997). The two faces of Adam: Ambivalent sexism and polarized attitudes toward women. *Personality and Social Psychology Bulletin, 23,* 1323-1334.

Glick, P. & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70,* 491-512.

Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychology of Women Quarterly, 21,* 119-135.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., & Masser, B. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology, 79,* 763-775.

Goldberg, L.R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheller (Ed.), *Review of personality and social psychology, Volume 2, (pp. 141-165).* Beverly Hills, CA: SAGE Publications.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Personality Processes and Individual Differences, 59,* 1216-1229.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26-34.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five-Factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.) *Personality psychology in Europe* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Goldman, S. H., & Raju, N. S. (1986). Recovery of one- and two-parameter logistic item parameters: An empirical study. *Educational and Psychological Measurement, 46,* 11-21.

Goldstein, H.W., Scherbaum, C.A., & Yusko, K.P. (2009). Revisiting g: Intelligence, adverse impact, and personnel selection. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 95-134). New York: Taylor & Francis.

Grijalva, E., Newman, D.A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). Gender differences in narcissism: A meta-analytic review. *Psychological Bulletin,* 141, 261-310.

Guyer, R. & Thompson, N. A. (2011). *User's manual for Xcalibre 4.1.* St. Paul, MN: Assessment Systems Corporation.

Guyer, R. & Thompson, N. A. (2014). *User's manual for Xcalibre item response theory calibration software, version 4.2.2 and later.* Woodbury MN: Assessment Systems

Corporation.

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

Hamilton, D.L., Gibbons, P.A., Stroessner, S.J., & Shermn, J. W. (1992). Stereotypes and language use. In G. R. Semin & K. Fiedler (Eds.), *Language, interaction, and social cognition* (pp. 102-128). London: Sage.

Hegelson, V.S. (2015). Gender and personality. In M. Mikulincer, P. R., Shaver, M. L., Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Vol 4. Personality processes and individual differences (pp. 515-534).* Washington, DC, US: American Psychological Association.

Hendriks, A. A. J., Hofstee, W. K. B., & De Raad, B. (1999). The Five-Factor Personality Inventory (FFPI). *Personality and Individual Differences, 27,* 307-325.

Hogan, R, & Chamorro-Premuzic, T. (2015). Personality and career success. In M. Mikulincer, P.R., Shaver, M. L., Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Vol 4. Personality processes and individual differences (pp. 619-638).* Washington, DC, US: American Psychological Association.

Horn, J. L. (1965). A rational and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.

Hough, L.M. & Connelly, B.S. (2013). Personality measurement and use in industrial and organizational psychology. In K. F. Geisinger, B.A. Bracken, J.F. Carlson, J. I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 501-531). Washington, DC: American Psychological

Association.

Hough, L.M., Ones, D.S., & Viswesvaran, C. (1998, April). Personality correlates of managerial performance constructs. In R. Page (Chair), *Personality determinants of managerial potential, performance progression and ascendancy.* Symposium conducted at 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Hough, L.M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9,* 152-194.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72-98.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581-592.

Jaffee, S. & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin, 105*, 198-214.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2,* 171-203.

John, O. P. & Gosling, S. D. (2000). Personality traits. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology* (Vol. 6, pp. 140-144). Washington, DC: American Psychological Association.

John, O.P., Robins, R.W., & Pervin, L.A. (Eds.) (2008). Handbook of personality: Theory and research (3rd ed.). New York, NY: Guilford Press.

John, O. P. & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and

theoretical perspectives. In L.A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research (*pp. 102-138). New York, NY: Guilford Press.

Johnson, M. A., Stevenson, R. M., & Letwin, C. R. (2018). A woman's place is in the… startup! Crowdfunder judgments, implicit bias, and the stereotype content model. *Journal of Business Venturing, 33,* 813-831.

Karanci, A. N., Dirik, G., & Yorulmaz, O. (2007). Reliability and validity studies of Turkish translation of Eysenck Personality Questionnaire revised – abbreviated. *Turkish Journal of Psychiatry, 18,* 1-7.

Kelly, G.A. (1955). The psychology of personal constructs. Vol. 1. A theory of personality. Vol 2. Clinical diagnosis and psychotherapy. Oxford, England: W.W. Norton.

Kervyn, N., Fiske, S., & Yzerbyt, V. (2015). Forecasting the primary dimension of social perception: Symbolic and realistic threats together predict warmth in the Stereotype Content Model. *Social Psychology, 46,* 36-45.

Kirisci, L., Hsu, T. C., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*, 146-162.

Kleinman, M. & Teresi, J. A. (2016). Differential item functioning magnitude and impact measures from item response theory models. *Psychological Test and Assessment Modeling, 58,* 79-98.

Landy, F.J. (1989). Psychology of work behavior (4th ed.). Belmont, CA: Thomson Brooks/Cole Publishing Co.

Larsen, R. J. & Buss, D. M. (2005). Introduction to personality. In *Personality psychology: Domains of knowledge about human nature* (2nd ed., pp 1-23). Boston: McGraw-Hill.

Linssen, H. & Hagendoorn, L. (1994). Social and geographical factors in the explanation of the content of European nationality stereotypes. *British Journal of Social Psychology, 33,* 165-182.

Lockenhoff, E., Chan, W., McCrae, R.R., De Fruyt, F., Jussim, L. De Bolle, M., …….. Terracciano, A. (2014). Gender stereotypes of personality: Universal and accurate? *Journal of Cross-Cultural Psychology, 45,* 675-694.

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum.

Lott, B. (1994). *Women's lives: Themes and variations in gender learning.* Pacific Giwe C.A: Books and Cole.

Lynn, R. & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 countries. *Journal of Social Psychology, 137,* 369-373.

Maass, A. & Arcuri, L. (1996). Language and stereotyping. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 193-226). New York: Guilford Press.

Maass, A., Montalcini, F., & Biciotti, E. (1998). On the (dis)-confirmability of stereotypic attitudes. *European Journal of Social Psychology, 28,* 383-402.

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

Marco, G. L. (1977). Item characteristic curve solution to three intractable testing problems. *Journal of Educational Measurement, 14,* 139-160.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits* (2nd ed.) Cambridge: Cambridge University Press.

McCrae, R. R. & Costa, P. T. (1996). Toward a new generation of personality theories:

Theoretical contexts for the five-factor model. In J.S. Wiggins (Ed.), The five-factor model of personality: Thereotical perspectives (pp. 51-87). New York, NY: Guilford Press.

McCrae. R. R. & Costa, P.T. Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509-516.

McCrae, R. R., & Costa, P. T. Jr. (1999). A Five-Factor theory of personality. In L. A. Pervin & O.P. John (Eds.) *Handbook of personality: Theory and research* (pp. 139-153). New York, NY: Guilford Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Messner, M. A. (1992). *Power at play: Sports and the problem of masculinity.* Boston: Beacon.

Mischel, W. (1996). *Personality and assessment.* Mahwah: Lawrence Erlbaum Associates.

Mitchelson, J. K., Wicher, E. W., LeBreton, J.M., Craig, S.B. (2009). Gender and ethnicity differences on the Abridged Big Five Circumplex (AB5C) of personality traits: A differential item functioning analysis. *Educational and Psychological Measurement, 69,* 613-635.

Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K., & Schmitt, M.J. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60,* 1029-1049.

Moskowitz, D. S. & Fournier, M. A. (2014). The interplay of persons and situations: Retrospect and prospect. In M. Mikulincer and P. R. Shaver (Eds.), *APA Handbook of Personality and Social Psychology, Vol. 4. Personality Processes and Individual Differences*

(pp. 471-489). Washington D.C.: American Psychological Association.

Nye, C. D., Allemand, M. Gosling, S. D., Potter, J., & Roberts, B. W. (2015). Personality trait differences between young and middle-aged adults: Measurement artifacts or actual trends? *Journal of Personality, 84,* 473-492.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60,* 995-1027.

Oswald, F.L. & Hough, L.M. (2011). Personality and its assessment in organizations: Theoretical and empirical developments. In S. Zedeck (Ed.), *APA Handbooks in Psychology. APA handbook of industrial and organizational, Vol 2. Selecting and developing members for the organization* (pp. 153-184). Washington, DC, US: American Psychological Association.

Patil, V. H., Singh, S. N., Mishra, S., & Donavan, T. (2007). Parallel analysis engine to aid determining number of factors to retain. Computer Software. https://analytics.gonzaga.edu/parallelengine/

Paunonen, S. V., & Hong, R. Y. (2015). *On the properties of personality traits.* In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook in psychology. APA handbook of personality and social psychology, Vol. 4. Personality processes and individual differences,* (p. 233-259). American Psychological Association.

Pervin, L. A. (1994). Personality stability, personality change, and the question of process. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 315-330). Washington, DC, US: American Psychological Association.

Pervin, L.A. (2001). A critical analysis of current trait theory. *Psychological Inquiry, 5,* 103-113.

Pew Research Center. (2013, Mar 2013). *Modern parenthood: Roles of moms and dads converge as they balance work and family.* [Report]. https://www.pewsocialtrends.org/2013/03/14/modern-parenthood-roles-of-moms-and-dads-converge-as-they-balance-work-and-family/

Philpot, C.L., Brooks, G. R., Lusterman, D.D., & Nut, R.L. (1997). *Bridging separate gender worlds: Why men and women clash and how therapists can bring them together.* Washington, DC: American Psychological Association.

Raju, N. (2000). Notes accompanying the differential functioning of items and tests (DFIT) computer program.

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

Rao, A. (2019). Even breadwinning wives don't get equality at home. *The Atlantic.* https://www.theatlantic.com/family/archive/2019/05/breadwinning-wives-gender-inequality/589237/

Rast, D. E., Gaffney, A. M., & Yang, F. (2017). The effect of stereotype content on intergroup uncertainty and interactions. *The Journal of Social Psychology, 158,* 711-720.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207-230.

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., & Suzuki, L. A. (2013). *Bias in psychological assessment: An empirical review and recommendations.* In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology: Assessment psychology* (p. 82-113). John Wiley & Sons, Inc.

Roivainen, E. (2013). Frequency of the uses of English personality adjectives: Implications for personality theory. *Journal of Research in Personality, 47,* 417-420.

Ross, C. E. & Van Willigen, M. (1997). Education and the subjective quality of life. *Journal of Health and Social Behavior, 38,* 275-297.

Rossier, J. (2015). Personality assessment and career interventions. In P.J. Hartung, M.L. Savickas, & W.B. Walsh (Eds.), *APA handbook of career intervention, Vol 1. Foundations* (pp. 327-350). Washington, DC, US: American Psychological Association.

Rothbart, M., & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology, 50,* 131-142.

Ruch W. (2014) Cheerfulness. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht

Rudman, L.A. & Phelan, J.E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior, 28,* 61-79.

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102,* 254-273.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34,* 100.

Saucier, G. & Goldberg, L.R. (1996). Evidence for the Big Five in analyses of familiar English personality adjectives. *European Journal of Personality, 10,* 61-77.

Saucier, G. & Srivastava, S. (2015). What makes a good structural model of personality? Evaluating the Big Five and alternatives. In M. Mikulincer and P. R.Shaver (Eds.), APA *Handbook of Personality and Social Psychology: Personality Processes and*

*Individual Differences, 4,* 283-305.

Scherbaum, C. (2003). *Detecting intentional response distortion on measures of the five-factor model of personality: An application of differential person functioning.* (Electronic Thesis or Dissertation). Retrived from https://etd.ohiolonk.edu/.

Scherwitz, L., Perkins, L., Chesney, M., & Hughes, G. (1991). Cook-Medley Hostility Scale and subsets: Relationship to demographic and psychosocial characteristics in young adults in the CARDIA study. *Psychosomatic Medicine, 53,* 36-49.

Schwarz, N. (1999). Self-reports: how the questions shape the answer. *American Psychologist, 54,* 93-105.

Sczesny, S., Nater, C., & Eagly, A. H. (2019). Agency and communion: Their implications for gender stereotypes and gender identities. In A. Abele & B. Wojciszke (Eds.), *Agency and communion in social psychology* (pp. 103-116). New York, NY: Routledge.

Sharp, C., Michonski, J., Steinberg, L., Fowler, J. C., Frueh, B. C., & Oldham, J. M. (2014). An investigation of differential item functioning across gender of BPD criteria. *Journal of Abnormal Psychology, 123,* 231-236.

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.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes. *Psychology of Women Quarterly, 24,* 44-62.

Spence, J. T. & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological*

*dimensions, correlates, and antecedents.* Austin: University of Texas Press.

Spence, J. T., Helmreich, R., & Stapp, J. (1973). A short version of the Attitudes toward Women Scale (AWS). *Bulletin of the Psychonomic Society, 2,* 219-220.

Soto, C. J., John, O.P., Gosling, S.D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100,* 330-348.

South, S. C., Reichborn-Kjennerud, T., Eaton, N. R., & Krueger, R. F. (2015). Genetics of personality. In M. In M. Mikulincer and P. R.Shaver (Eds.), APA *Handbook of Personality and Social Psychology: Personality Processes and Individual Differences, 4*, 31-60.

Stake, J.E., & Eisele, H. (2010). Gender and personality. In J.C. Chrisler & D. R. McCreary (Eds.), *Handbook of gender research in psychology, Vol 2. Gender research in social and applied psychology* (pp. 19-40). New York, NY: Springer Science + Business Media.

Stangor, C. & Lange, J.E. (1994). Mental representations of social groups: Advances in understanding stereotypes and stereotyping. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 357-416). San Diego, CA: Academic Press.

Stangor, C. & Schaller, M. (2000). Stereotypes as individual and collective representations. In C. Stangor (Ed.), *Key readings in social psychology. Stereotypes and prejudice: Essential readings* (pp. 64-82). New York, NY: Psychology Press.

Tajfel, H. & Turner, C. J. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations.* Chicago, IL: Nelson-Hall.

Thissen, D. & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika,*

*47*, 397-412.

Thurstone, L.L. (1934). The vectors of mind. *Psychological Review, 41*, 1-32.

Tupes, E. C. & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60,* 225-251.

Uher, J. (2013). Personality psychology: Lexical approaches, assessment methods, and trait concepts reveal only half of the story. Why it is time for a paradigm shift. *Integrative Psychological and Behavioral Science, 47,* 1-55.

Vianello, M. Schnabel, K., Sriram, N., & Nosek, B. (2013). Gender differences in implicit and explicit personality traits. *Personality and Individual Differences, 55,* 994-999.

Wang, W.C. (2008). Assessment of differential item functioning. *Journal of Applied Measurement, 9,* 387-408.

Wasserman, J. D., & Tulsky, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp.3-22). New York, NY: Guilford Press.

Wetzel, E., Bohnke, J. R., Carstensen, C. H., Ziegler, M., & Ostendorf, F. (2013). Do individuals response styles matter? Assessing differential item functioning for men and women in the NEO-PI-R. *Journal of Individual Difference, 34,* 69-81.

Winter, D.G. & Barenbaum, N.B. (1999). History of modern personality theory and research. In L.A. Pervin & O.P. John (Eds.), *Handbook of personality: Theory and research* (pp. 3-27). New York, NY, US: Guilford Press.