USING MM-IRT-C TO EXPLORE THE RELATIONSHIP BETWEEN DEPRESSION AND PRE-EMPLOYMENT TESTS

Rachel T. King

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2017

Committee:
Michael Zickar, Advisor
Virginia L. Dubasik
Graduate Faculty Representative
Russell Matthews
John Tisak
This research examines personality test response patterns between individuals with depression and without depression, and how the relationship between depression and employment testing relates to job performance. MM-IRT-C was used to analyze the Big Five for latent classes, and to determine if the latent classes that emerged could be explained by depression and social desirability as covariates. Results showed that social desirability was an important covariate in explaining the latent classes that emerged across the Big Five, but that depression was not a significant covariate for most personality constructs. Details of the latent class analyses and implications are discussed.
To Tom and my parents, David and Marina - Thank you for all of your support, encouragement, and sacrifice.
ACKNOWLEDGMENTS

I’d like to thank my advisor, Mike Zickar for being such an incredible mentor, and for all of his feedback during the dissertation process. I’d also like to thank my committee for their insightful comments and feedback on this manuscript. Finally, I’d like to thank my graduate school colleagues and my family for their support and friendship during this process.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Personality</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Personality Testing</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>The Dark Triad</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Personality and Mental Health</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Personality and Depression</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>MM-IRT</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Job Performance</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Job Performance and Depression</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>2. HYPOTHESES</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>3. METHOD</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Participants</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Measures</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Demographics</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Personality</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Social Desirability</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Clinical Diagnoses</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Task Performance</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>OCB</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>CWB</td>
<td>25</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>4</td>
<td>RESULTS</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Descriptive Analyses</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Invariance Analyses and Mean Differences</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>MM-IRT-C</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>DISCUSSION</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Practical Implications</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Limitations and Future Directions</td>
<td>44</td>
</tr>
</tbody>
</table>

REFERENCES ................................................................. 47

APPENDIX A. INSTRUMENT ..................................................... 54

APPENDIX B. HSRB LETTER .................................................... 59
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Means, Standard Deviations, Reliabilities, and Inter-scale Correlations (n = 947 - 1026)</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>MM-IRT-C Models (Final Model for each construct is italicized)</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>Correlations Between Personality Latent Classes and the Criterion (n = 947 - 1026)</td>
<td>62</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean Response Pattern for Openness classes, y-axis - probability of endorsement; x-axis - scale items</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>Mean Response Pattern for Agreeableness classes, y-axis - probability of endorsement; x-axis - scale items</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>Mean Response Pattern for Neuroticism classes, y-axis - probability of endorsement; x-axis - scale items</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>Mean Response Pattern for Conscientiousness classes, y-axis - probability of endorsement; x-axis - scale items</td>
<td>66</td>
</tr>
<tr>
<td>5</td>
<td>Mean Response Pattern for Extraversion classes, y-axis - probability of endorsement; x-axis - scale items</td>
<td>67</td>
</tr>
</tbody>
</table>
CHAPTER 1. INTRODUCTION

There have been recent articles in the popular press and within the field of industrial/organizational psychology about the potential for employment tests to be biased against individuals with personality disorders or mental illnesses such as depression (e.g. Below, 2014; Weber & Dwonskin, 2014). This potential for bias is problematic because mental health status is protected under the American with Disabilities Act, and it is illegal to discriminate against individuals with mental illness. CVS settled a civil case out of court in 2011 brought by the Rhode Island American Civil Liberties Union that claimed items in their personality testing battery were biased against individuals with mental illnesses, and there have been recent investigations into the personality tests being used by Kroger and RadioShack (Weber & Dwonskin, 2014). This argument against personality testing is that individuals who suffer from mental health conditions respond to personality inventories in a different way than individuals who do not have a diagnosable condition. This, in turn, could potentially lead to these individuals not getting jobs due to their mental health status.

In this research, I examined how the results from employment tests can be used to categorize individuals with depression, a common mental health concern, and how the relationship between depression and employment testing relates to job performance. I examined if personality scales that are often used for hiring purposes function differently for individuals who have been diagnosed with depression or score highly on a commonly used diagnostic scale, compared to those who do not score highly on these measures or have not been diagnosed. Depression was selected because it is the most common mental illness; according to the most recent figures from the National Survey on Drug Use and Health it affected approximately 6.9%
of the US population over the age of 18 in 2012 (Substance Abuse and Mental Health Services Administration, 2013).

I used a Mixed-Measure Item Response Theory model with covariates (MM-IRT-C) to examine this empirically with measures of the Big Five. The benefit of using MM-IRT-C techniques over other possible methodologies was that this technique allowed me to examine how individuals respond to these scales at an item level and to determine if there were latent classes in the response patterns. Once I identified latent classes in the personality scales using MM-IRT-C techniques, I examined if there are differences in how the individuals in these various latent classes perform on the job. I used organizational citizenship behaviors (OCB), counterproductive work behaviors (CWB), and task performance as my criterion measures, and I tested how well latent class membership predicts performance. I also checked if there were differences in performance among groups. This project allowed for an investigation of how personality scales function among those with and without mental health issues and how this relationship between personality and depression impacts the relationship between personality and criteria.

It is important to understand how personality scales function across groups because they are frequently used as an applicant screening tool in making employment decisions. Scales that function differently across groups can be biased if the groups respond to the same items in different ways. For example, an item on an intelligence test that references football may receive very different responses from an American respondent who thinks the item is referring to American football than a European respondent who believes the item refers to soccer. The difference in responses is not due to a true difference in the levels of intelligence between the two respondents, but is a characteristic of the item that is asked. It should be pointed out that
group differences can also occur when there truly are mean differences between two groups on an underlying construct; an examination of item bias can help tease apart whether group differences are genuine or if they are a by-product of how the two groups use the scale. If personality scales are biased against individuals with depression, then steps can be taken to reduce this bias. This can be accomplished in several different ways. First, if organizations are aware of bias in a commonly used scale, they can be more cautious in how it is used. Measures of intelligence are known for showing strong racial differences (Sackett, Schmitt, Ellingson, & Kabin, 2001) and accordingly, are used judiciously in employee selection. Personality scores are typically fairly consistent across race and sex. The second way bias can be reduced is by removing problematic items that are driving the discrepancy between groups. If there are just a few items that are driving the group difference, for example, neuroticism items that tap into depression, it may be possible to remove these items from a scale and reduce or remove the group differences. In summary, this psychometric research is important because it can help bring unfairness to light and potentially shed some light on how to reduce any bias in personality measurement.

Taking a step back, it is also important to consider mean differences in how applicants with and without mental illnesses respond to personality inventories. Mean differences between protected classes, such as individuals suffering from depression, and non-protected classes are allowed if a test can be shown to be job relevant, and if it can be shown that there is not an equally valid test with a lower level of adverse impact (Equal Employment Opportunity Commission (EEOC), Civil Service Commission, Department of Labor, & Department of Justice, 1978). However, this is not an ideal situation as it can still open an organization up to lawsuits and increased scrutiny of their selection practices. Should the personality scales used
demonstrate scalar invariance, which indicates that the scales are measuring the same underlying construct in both individuals with and without mental illness (Vandenberg & Lance, 2000), I will also examine mean differences between individuals with and without mental illness. It is important to note that if a scale does not demonstrate scalar invariance, then the means of two groups cannot be reasonably compared.

In this paper, I begin with a discussion of how non-pathological personality has been assessed, then I review how personality assessment is used in employment screening. From there I move into a discussion of dark personality traits followed by a discussion of pathological personality and the relationship between personality and depression. The next section outlines the mixed model item response theory techniques that are used for the data analysis, followed by a discussion of job performance and its relationship with depression. The hypotheses and methods are then followed by the results and a discussion of the results.

**Personality**

The discussion of personality in this section begins with a general overview of non-pathologic personality, then this section moves into a discussion of how personality testing is currently used in employment testing. This section next outlines the dark side of personality which can be seen as being on the more extreme side of non-pathologic personality, and then moves into a general discussion of personality and mental health. The section concludes with discussion of personality and depression.

Personality is a trait level construct that describes patterns in human behavior, and it is a commonly measured construct in the field of I/O psychology. Personality measures have been shown to be valid predictors of employee performance (Barrick & Mount, 1991), and can reduce
the adverse impact, group differences that lead to inequitable hiring decisions between a protected and non-protected group, in a selection system when they are used in conjunction with cognitive ability compared to using cognitive ability tests alone (De Corte, Lievens, & Sackett, 2007). Although there are many different models of personality, the Big Five, a commonly used personality model, has been the subject of much research within our field, is empirically supported (Barrick & Mount, 1991), and is the basis for some of our most commonly used measures (e.g. the International Personality Item Pool (IPIP) and the NEO Personality Inventory (NEO-PI). The Big Five divides personality into five factors Extraversion-Introversion, Conscientiousness, Openness to Experience, Neuroticism-Emotional Stability, and Agreeableness. Each of these factors of personality is briefly summarized below.

*Extraversion-Introversion* can be thought of as a continuum from highly outgoing and needing a great deal of social interaction, to less outgoing and requiring less social interaction. An extraverted individual will tend to be more talkative, more socially dominant, and more outgoing than an introverted individual. Some researchers have suggested that this facet actually contains two separate sub-dimensions, sociability and surgency (Hogan, 1992). This view separates the component of extraversion that involves how much one enjoys socializing and being the center of attention from the component that deals more with social dominance. This facet of personality as a whole has been shown to predict performance for jobs where social interaction is required and is commonly used as a selection test for jobs of this nature (e.g. sales; Barrick & Mount, 1991).

An individual who is high on the dimension of *Conscientiousness* will tend to be punctual, dependable, organized, and ensure that things are done the correct way. Conscientiousness is related to many important outcomes in both the work and the personal domains. For example,
Conscientiousness is associated with engaging in positive health behaviors such as screenings and following through on treatment protocols, and avoiding more risky behaviors such as drug and alcohol use (Bogg & Roberts, 2004). Conscientiousness is also related to important work outcomes. Conscientiousness has been shown to predict job performance regardless of the job type (Barrick & Mount, 1991; Schmidt & Hunter, 1998), and because of this, is the most commonly assessed dimension of personality within a selection context.

*Openness to Experience*, also referred to as intellectance or just openness, is the most debated dimension of the Big Five, in that it is difficult to identify and changes slightly cross-culturally. This dimension is characterized by a willingness to try new things, being open-minded to new ideas, being more liberal, and being creative. Someone who is lower on this dimension will be more set in their ways and less likely to try new things. This dimension is typically unrelated to job performance for most jobs that do not involve creative work and is unrelated to mental illness (Barrick & Mount, 1991; Malouff, Thorsteinsson, & Schutte, 2005).

*Neuroticism-Emotional Stability* is a dimension characterized by high levels of emotionality, moodiness, and rumination at the neuroticism end of the spectrum. A person who is closer to the emotional stability end of the spectrum could be characterized as even-keeled or laid back. An emotionally stable person is more in control of his or her emotions than someone who falls on the neuroticism side. This dimension has also been used in selection research and has been found to predict performance for many jobs (Barrick & Mount, 1991). This dimension is also highly correlated with mental illness as will be discussed later in greater detail.

Finally, *Agreeableness* is characterized by friendliness, a willingness to cooperate and compromise with others, and a willingness to help others. Someone who is agreeable will also be
collaborative and will get along well with others. A lack of agreeableness can be characterized by stubbornness, a willingness to argue or fight, and surliness. This dimension of personality, like extraversion, can be an excellent predictor of performance for jobs where agreeableness is relevant to the tasks required by the job such as sales and customer service positions (Barrick & Mount, 1991).

In addition to the five factors, subfacets for each of these personality factors have been proposed (Costa & McCrae, 1995). Subfacets are more specific personality traits that load onto one of the Big Five. For example, conscientiousness is comprised of six lower-order facets including self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, and cautiousness. Subfacets are measured for a variety of reasons. When predicting narrow job outcomes, it can be beneficial to match the bandwidth of the predictor to the criteria (Hogan & Roberts, 1996). For example, if a hiring manager is seeking to hire an assistant who will be very meticulous and keep all of the manager’s appointments well organized, the subfacet of orderliness may be a job relevant predictor for the assistant position. Subfacets can also help ensure that the full range of a construct is being captured.

These five dimensions and their subfacets comprise personality in the five-factor model, and their relationships with mental health and job performance will be discussed. Now that we have established what personality is, it is important to understand how it is measured and why.

**Personality Testing.** Personality tests have been the subject of a fair amount of research within I/O psychology and are a popular form of pre-employment screening. Personality tests are valid predictors of performance (Barrick & Mount, 1991) and demonstrate less adverse impact than cognitive ability tests. Personality tests that are developed for employment testing typically
focus on the non-pathologic side of personality and are not designed for clinical diagnoses. Tests that explicitly measure psychopathology such as the Minnesota Multiphasic Personality Inventory (MMPI) are not used for employment screening because they are designed to detect mental illness and are considered a medical exam under the American with Disabilities Act and are expressly forbidden as a pre-employment screening expect for certain positions involving public safety, such as the hiring of police officers.

Recently, employment tests that use personality have received attention in the media because there have been several lawsuits claiming that personality tests are biased against those with diagnosed mental illnesses even if these tests are not explicitly designed to measure mental illness (Weber & Dwonskin, 2014). Among these lawsuits was an instance in which CVS settled a case claiming that their personality tests violated state laws prohibiting questions about mental health prior on employment screening tools out of court in 2011. Kroger is also facing ongoing investigations by the EEOC into their personality tests and whether or not these tests discriminate against those with mental illness. The article by Weber and Dwonskin (2014) also suggests that the EEOC has an ongoing investigation into personality testing as pre-employment screen in general stating, “The Equal Employment Opportunity commission is investigating whether personality tests discriminate against people with disabilities. As part of the investigation, officials are trying to determine if the tests shut out people suffering from mental illnesses such as depression or bipolar disorder, even if they have the right skills for the job, according to EEOC documents.” However, the article does not provide any more concrete information regarding the investigation, claiming the EEOC is unwilling to discuss the ongoing investigation. This article does not discuss the difference between personality tests developed for mental health assessment, which are considered a medical screening device and cannot be used
as precondition for employment, and personality tests that are designed to measure non-clinical personality traits. The article also points to specific questions within a test as examples of questions that could be discriminatory, but fails to realize that hiring decisions are not made on a single question alone.

Although the information presented in these media reports presents a somewhat flawed view of employment testing by implying that decisions can be made based on a single item within an inventory or that the item level responses are available to managers, it does raise some interesting questions about how individuals with and without mental health issues respond to personality inventories and if they are responding in the same manner as one another. Even though decisions cannot be made on a single question, if there are enough questions that function differently for those with and without mental illness, this can lead to a test that lacks invariance overall.

The question at hand is whether the types of personality tests that are often administered prior to employment discriminate against those with mental illness. Because this is a very broad question, I have focused specifically on the most common mental illness in the US population, depression. To better address this question it is first important to understand the darker side of human personality.

**The Dark Triad.** Although they are not the focus of this research, it is important to understand the dark personality traits that lie between normal personality and pathology. The Dark Triad are sub-clinical traits that exist in the grey area between personality and mental illness. Although they are not diagnosable mental illnesses, there has been a fair amount of research in I/O dealing with the *Dark Triad* of personality and its relationship with performance.
Although the *Dark Triad* are not clinical diagnoses, they are typically thought of as being on the extreme end of non-pathological personality. The *Dark Triad* of personality consists of Machiavellianism, narcissism, and psychopathy (O'Boyle, Forsyth, Banks, & McDaniel, 2012). Narcissism is characterized by an extreme self-interest and having an inflated view of oneself. Machiavellianism is comprised of three facets: manipulating others, a willingness to cut corners and bend rules to accomplish goals, and cynicism in one’s view of human nature. Finally, psychopathy is characterized by a disregard for others and a lack of regret when others are harmed due to one’s actions.

O’Boyle and colleagues (2012) conducted a meta-analysis of the relationship between the *Dark Triad* and performance and found that Machiavellianism and psychopathy are negatively related to job performance, and that all three personality traits are positively related to counter-productive work behaviors (CWB). This suggests that negative personality traits may have an impact on job performance and should be studied further. Although these three personality traits are not necessarily pathologic, they are closer to the pathologic end of the spectrum than many other commonly measured personality traits such as the Big Five. There has been some research that suggests that personality is a continuum and pathologic and non-pathologic traits are on the same spectrum (Dilchert, Ones, & Krueger, 2014); this can make it difficult to draw a clear line between a diagnosable mental illness that is protected under federal law and an undesirable personality trait that hinders job performance and is fair game for a pre-employment screening test. One common way to differentiate between clinical and non-clinical personality traits is whether or not the trait causes the individual distress and/or if it hinders the individual's daily life. Someone who is mildly narcissistic, for example, may not be a very good team player, but may still be able to lead a normal life. However, if that person's narcissistic tendencies are so extreme
as to prevent him or her from obtaining gainful employment and from having healthy
relationships with friends and family members, that personality trait crosses the line into a
personality disorder. Moving further down the continuum between pathologic and non-
pathologic personality, we turn now to personality and mental health.

**Personality and Mental Health.** Although I focused on the relationship between
personality and depression, it is important to understand the broader context in which research on
personality and depression is set. First of all, it is important to understand the relationship
between personality and mental illness. Secondly, many of the meta-analyses and generalized
findings reported in this section inform the literature on personality and depression and provide
additional evidence for a relationship, or lack thereof, between depression and various
personality traits. In this section, I will discuss the broad relationship between personality and
Axis I clinical disorders (American Psychiatric Association, 2013). Then I will move into a
discussion of the relationship between personality and Axis II personality disorders (American
Psychiatric Association, 2013), and finally end with a more in depth discussion of the
relationship between personality and depression.

There has been a fair amount of research looking at the relationship between personality
and common clinical disorders, such as anxiety disorders, substance use disorders, and
depressive disorders. Several meta-analyses have been conducted on the topic. Malouff,
Thorsteinsson, and Schutte (2005) found a meta-analytic relationship between mental illness and
the Big Five where most disorders were correlated with low conscientiousness, extraversion, and
agreeableness. They also found that most disorders were correlated with high neuroticism, and
were uncorrelated with openness. Kotov and colleagues (2010) used a broader sample and
explored both the Big Five and the Big Three and also found evidence for correlations between
mental illnesses and high level of neuroticism and low levels of conscientiousness. They also found relationships between substance use disorders and low levels of agreeableness, and found that some disorders, including depression, were negatively related to extraversion. Like Malouff and colleagues (2005), Kotov et al. (2010) did not find strong relationships between mental illness and openness to experience and neither paper found evidence for a relationship between agreeableness and depression.

Traditional personality scales are also used to help inform diagnoses and treatment options in clinical settings (Costa & McCrae, 1992). Treatment options can be made more accurately by taking into account a patient’s personality. For example, a more extraverted patient may respond better to a therapy method that requires talking about oneself than a more introverted patient would. In addition to aiding in treatment decisions, personality can also help diagnose certain conditions. Neuroticism can be used to help inform diagnoses of anxiety and depression. One of the differentiators between a diagnosis of anxiety and a diagnosis of depression is how neurotic a patient is (Costa & McCrae, 1992).

In addition to the research on common clinical disorders discussed above, there has also been research on the relationship between personality and personality disorders such as research correlating the Dimensional Assessment of Personality Pathology and the Big Five (e.g. Schroeder, et al. 1992). Personality disorders are less common than many of the clinical disorders discussed above. The authors found that personality disorders correlate with the Big Five, especially with neuroticism sub-facets. Various indicators of personality disorders loaded onto all of the Big Five facets with the exception of openness to experience. Whereas most of the relationships between personality and personality disorders are negative, those with compulsive tendencies tend to score higher on measures of conscientiousness than the general population.
Even though compulsive tendencies may detract from job performance, if individuals with compulsive tendencies have higher scores on measures of conscientiousness they may be more likely to be selected into the organization than those with other mental health conditions.

**Personality and Depression.** Turning more specifically to depression, there has been a fair amount of research on how personality relates to depression and how depression can impact the measurement of personality. There are a variety of classical models and theories regarding the relationship between personality and depression. These generally fall into one of three types of models, models where personality and depression share a third variable that is driving both of them or are ends of the same underlying continuum (e.g., models where personality has a causal influence on depression, and models where depression has a causal influence on personality; Klein, Kotov, & Bufferd, 2011). More recently, dynamic models have begun to emerge in the literature that posit a reciprocal relationship between depression and personality. These dynamic models acknowledge that there can be changes in personality throughout the lifetime (Roberts, Walton, & Viechtbauer, 2006) and allow for the temporary altering of personality by depressive symptoms (Klein et al., 2011). All of these theoretical models acknowledge a relationship between certain personality traits and depression, though they disagree on the causal direction and the presence of outside drivers.

Depression has been linked to higher levels of self-reported neuroticism, such that during a depressive episode individuals report higher levels of neuroticism than reported by the same individuals when they are not actively suffering from a depressive episode (Hirschfeld, Klerman, Clayton, Keller, McDonald-Scott, & Larkin, 1983; Liebowitz, Stallone, Dunner, & Fieve, 1979). Additionally, Kotov et al. (2010) found a strong relationship between neuroticism and depression in their meta-analysis.
Based on these findings regarding the relationship between the Big Five and depression, I expected to find that neuroticism scores would be more likely to be impacted by depression status than scores on the other four factors will be impacted by depression. In addition to the evidence cited above, when looking at the facet level of personality, one of the commonly assessed subfacets of neuroticism is labeled depression. Although this is not the same as a clinical diagnosis, there are similarities in the item content between a scale such as the subfacet of depression for neuroticism in the NEO-IPIP (Goldberg, et al., 2006) and the items on the CES-D depression inventory (Radloff, 1977). Because of this overlap in the operationalization of neuroticism and clinical depression scales, latent classes may emerge due to overlap in items and due to the conceptual overlap between the constructs. Additionally, research at the subfacet level has found that individuals with major depressive disorder have different means on the subfacets of neuroticism and the size of mean difference between the depressed and non-depressed groups varies across the subfacets (Bienvenu, Samuels, Costa, Reti, Eaton, & Nestadt, 2004). This is important for latent class membership because an overall measure of neuroticism will have items that capture aspects of various subfacets, leading to different response patterns across items. If the pattern of responses changes at the subfacet level, then this suggests that the pattern of responses will also change at the facet level.

Turning to other personality factors, individuals suffering from a depressive episode have been found to score lower on measures of conscientiousness than the general population, and have been found to still have lower average levels of conscientiousness up to six months after a depressive episode (Anderson & McLean, 1997; Kotov et al., 2010). Individuals suffering from depression have also been found to display lower levels of extraversion (Jylhä, & Isometsä, 2006; Kotov et al., 2010; Malouff et al., 2005). Conscientiousness and extraversion also showed
different means at the subfacet level between depressed and non-depressed individuals across the subfacets (Bienvenu et al., 2004). I expected to find latent classes emerging on both the conscientiousness and extraversion scales in addition to finding mean differences. This is important because these differences would indicate that individuals with depression are scoring lower on these scales than individuals without depression and that the presence of depression is driving these differences.

**MM-IRT**

As I have discussed, more recent research suggests that personality disorders and non-pathologic personality are on the same spectrum, but because they are so far removed from one another they are assessed with separate inventories (Dilchert, Ones, & Krueger, 2014). This suggests that some items on personality assessments may function differently for those with personality disorders, particularly when personality is measured at the facet level as opposed to the subfacet level. For example, items on a neuroticism scale that are most closely connected to the neuroticism subfacet *depression* may show greater differences between a depressed and non-depressed population than an item from the neuroticism subfacet of *impulsivity*. How an inventory behaves between groups can be examined using mixed model item response theory.

Mixed model item response theory (MM-IRT) is a statistical technique that combines latent class analysis with item response theory. This technique allows researchers to determine if there are latent classes of respondents to the items within a scale and determine if these latent groups of respondents respond differently from one another. The grouping is accomplished by identifying groups of respondents who have similar item response functions, and grouping those with similar response patterns into latent classes.
This technique provides several benefits over a traditional differential item functioning (DIF) analysis. In a traditional DIF analysis the grouping variable must be dichotomous. When dealing with constructs that are continuous such as scores on a Likert scale or age variables the scales must be artificially dichotomized in order to conduct a DIF analysis, which can decrease power to detect an effect and oversimplify conclusions that are reached from the data analysis. Because this study will be using a continuous measure of social desirability, artificially dichotomizing the scale is not desirable. With MM-IRT continuous covariates can be analyzed as they are. Another benefit is that groups do not have to be established a priori for a MM-IRT as they do for a DIF analysis. MM-IRT-C (Tay, Newman, & Vermunt, 2011) allows covariates to be included directly in the model, meaning that depression and can be added directly into the model as a covariate this will provide more accurate information about whether the DIF is originating from the depression variable or from another source. Having covariates in the model to explain the latent classes also allows for hypothesis testing, which cannot be done directly within a DIF analysis. To run an MM-IRT model the researcher first fits a model with one latent class, then two classes and so on until there is not an improvement in model fit from the addition of an extra latent class (Carter, Dalal, Lake, Lin, & Zickar, 2011). MM-IRT can be conducted with a variety of available software programs including Latent GOLD and MDLTM.

Past research using the MM-IRT technique has examined a variety of topics including latent classes in response patterns to the Job Descriptive Index (Carter et al., 2011). Carter and colleagues found that there were three classes of respondents to the Job Descriptive Index, a measure of job satisfaction, those most likely to say “yes,” those most likely to say “no,” and those most likely to use the “?” option. Hernández, Drasgow, and González-Romá (2004) found
that there are two different subpopulations in terms of usage of the middle response option on Likert scales.

Research using MM-IRT has also been conducted on various personality scales. Zickar, Gibby, and Robie (2004) used MM-IRT techniques to model faking behavior on the ABLE personality instrument used by the US military and found two classes of fakers along with an honest class. Another paper examining personality using MM-IRT techniques attempted to improve prediction by analyzing latent classes of personality separately from another (Maij-de Meij, Kelderman, & van der Flier, 2008). The authors analyzed the personality scales from the Amsterdam Biographical Questionnaire and found that latent classes emerged based on both social desirability and the participant’s ethnic background. Because their data was collected in the Netherlands, the mix of ethnic backgrounds in their sample was somewhat unique and more varied than the data that can be obtained through an Mturk sample. Therefore, I did not expect ethnicity to be major determinant of the latent classes that emerge, but socially desirable responding was expected to play a role in the classes that emerged based on the research of Maij-de Meij and colleagues (2008) and Zickar and colleagues (2004).

**Job Performance**

As was previously discussed, personality tests predict job performance (Barrick & Mount, 1991), although the relevant facets may vary by job type. A broad view of the criterion will be taken for this study. Performance will be operationalized as consisting of task performance, organizational citizenship behaviors, and counterproductive work behaviors.

Task performance can be defined as the component of job performance that is directly related to one’s tasks and duties (Viswesvaran & Ones, 2000). Task performance is typically rated by an employee’s supervisor. Both subjective measures of task performance (e.g. holistic
supervisor ratings) and objective measures of task performance (e.g. sales figures for the quarter) can be used to assess task performance. Because of the limitations of collecting a large sample on MTurk, this study will use self-rated subjective performance. There is a precedent for using self-report performance data in studies that examine the impact of depression on job performance. The research on presenteeism that is discussed in the next section relies almost exclusively on self-report data.

Organizational Citizenship behaviors (OCB) are the component of job performance that deals with extra-role behaviors (Podsakoff, MacKenzie, Paine, & Bachrach, 2000). This includes behaviors that benefit one’s coworkers, such as helping a coworker finish a project or decorating someone’s office on his or her birthday (OCB-I). It also includes behaviors that benefit the organization such as being a team player and being loyal to the organization (OCB-O).

Counterproductive work behaviors (CWB) can be conceptualized as any behavior intended to hurt the organization or someone within the organization (Sackett & DeVore, 2001). CWBs aimed at others in the organization (CWB-I) can range from minor incivility, which is low-level deviance with an ambiguous motive, such as not saying hello to someone in the hall, to more extreme forms of deviance such as violence in the workplace. CWBs targeting the organization (CWB-O) include behaviors such as time theft, stealing office supplies, calling in sick when one is not actually sick, and working slowly.

These various indicators of performance are included partially to ensure that full domain of the criteria is covered, and partially based on past research that has found differences in relationships between personality and the criteria depending on the criteria measure used such as the research of O’Boyle and colleagues (2012) who found that different dark personality traits predicted counterproductive work behaviors and task performance.
**Job Performance and Depression.** There has also been a history of job performance and depression research particularly within the areas of vocational and clinical psychology in addition to a few epidemiological studies. Some of these studies that are outlined in this section do not define job performance in the same manner as it is typically defined within the area of industrial organizational psychology; however, taken with the research within our field, it helps to paint a picture of how depression can lead to performance deficits on the job.

In a review of the depression and job performance literature, Lerner and Henke (2008) found that individuals with depression are more likely to be unemployed or underemployed, are more likely to be absent from work, and are more likely to suffer from presenteeism in the form of lower levels of job performance and decreased productivity at work. For example, a psychiatric study showed that patients with depression showed greater performance deficits as measured by the Work Limitations Questionnaire, a measure of presenteeism, when compared to a group of patients with rheumatoid arthritis. The patients with depression were shown to have more difficulty with physical tasks, time management, and output tasks compared to the control group (Adler, McLaughlin, Rogers, Chang, Lapitsky, & Lerner, 2006). Presenteeism has also been studied using experience sample methods. Wang and colleagues (2004) conducted an experience sampling study in which participants reported on their job performance and productivity throughout the day. The study included several chronic conditions including headaches, back pain, arthritis, depression, and others. Depression was the only condition that showed significant effects of the condition on presenteeism.
CHAPTER 2. HYPOTHESES

Although most of the meta-analyses that were presented focused on mental illness at a broader level (Kotov et al., 2010; Malouff et al., 2005), they still provide compelling evidence for a lack of relationship between openness to experience and depression as well as a lack of a relationship between agreeableness and depression.

Hypothesis 1. Openness to experience will not show latent classes that can be explained by depression as a covariate.

Hypothesis 2: Openness to experience will not be correlated with depression.

Hypothesis 3. Agreeableness will not show latent classes that can be explained by depression as a covariate.

Hypothesis 4: Agreeableness will not be correlated with depression.

Turning to the relationship between depression and the other four facets of personality, based on prior research I expected that neuroticism would be positively related to depression (Hirschfeld, Klerman, Clayton, Keller, McDonald-Scott, & Larkin, 1983; Kotov et al., 2010; Liebowitz, Stallone, Dunner, & Fieve, 1979), conscientiousness would be negatively related to depression (Anderson & McLean, 1997; Kotov et al., 2010) and depression would be negatively related to extraversion (Jylhä, & Isometsä, 2006; Kotov et al., 2010; Malouff et al., 2005).

Hypothesis 5. Neuroticism will be correlated with depression, such that those with higher scores on the depression measure will also report higher scores on the measure of neuroticism.
Hypothesis 6. Neuroticism will show latent classes that can be explained by depression as a covariate such individuals with depression will be more likely to have higher item level scores on the measure of neuroticism.

Hypothesis 7. Conscientiousness will be negatively correlated with depression, such that those with higher scores on the depression measure will also report lower scores on the measure of conscientiousness.

Hypothesis 8. Latent classes will emerge on the conscientiousness scale based on whether or not an individual suffers from depression such that individuals with depression will be more likely to have lower item level scores on the measure of conscientiousness.

Hypothesis 9. Extraversion will be negatively correlated with depression, such that those with higher scores on the depression measure will also report lower scores on the measure of extraversion.

Hypothesis 10. Latent classes will emerge on the extraversion scale based on whether or not an individual suffers from depression such that individuals with depression will be more likely to have lower item level scores on the measure of extraversion.

In addition to seeing latent classes emerged based on depression as a covariate, I also expected socially desirable responding to play a role in the classes that emerge based on the research of Maij-de Meij and colleagues (2008) and Zickar and colleagues (2004).

Hypothesis 11. The latent classes will be at least partially explained by socially desirable responding.
Finally, I expected that the latent classes derived from the MM-IRT-C model would
differentially predict job performance based on past research that found mental illness, including
depression had a negative impact on productivity and career outcomes (e.g. Ettner, Maclean, &
French, 2011; Lim, Sanderson, & Andrews, 2000). I also expected to find a main effect such that
those with higher depression scores would have lower self-reported task performance.

_Hypothesis 12a._ The latent classes derived from the MM-IRT-C analysis will
differentially predict task performance such that those in latent classes
characterized by depression will have lower scores on task performance.

_Hypothesis 12b._ The latent classes derived from the MM-IRT-C analysis will
differentially predict OCB such that those in latent classes characterized by
depression will have lower scores on OCB.

_Hypothesis 12c._ The latent classes derived from the MM-IRT-C analysis will
differentially predict CWB such that those in latent classes characterized by
depression will have higher scores on CWB.

_Hypothesis 13._ Depression scores will be correlated with task performance, such
that those with higher scores on the depression measure will also report lower
scores on the measure of task performance.
CHAPTER 3. METHOD

Participants

Data for this study was gathered using a sample from MTurk. Previous research using MM-IRT techniques have had samples of between 1,000 and 6,000 (e.g. Tay et al., 2011, Zickar et al., 2004). For this project data the target sample size was 1,000 participants; this was partially to ensure that the sample overall was large enough to detect an effect if an effect is there and also to ensure that there were a sufficient number of individuals in the sample who have higher scores on the measures of psychopathology.

MTurk was a suitable source of data for this project for two reasons. First of all, depression has a higher rate of prevalence in the U.S. population than many other common clinical conditions (Substance Abuse and Mental Health Services Administration, 2013). Secondly, because this investigation is focused on the labor force and the selection process, using a clinical population may result in the inclusion of participants with more severe conditions who are not in the labor force. MTurk has been found to be an acceptable source of high quality data when proper best practices, such as including attention checks, only allowing workers with high approval ratings, and only allowing workers from the country of interest to participate are followed (Highhouse & Zhang, 2015; Landers & Behrend, 2015). Seven attention checks were used ranging from nonsense items (e.g. [I] am a toy poodle.) to items that instructed the participants to select a specific response options (e.g. Choose the middle response.).

Measures

Demographics. For this project data was collected from 1,164 participants through MTurk. Participants who missed more than two attention checks were removed from the sample leaving a final sample of 1,026 participants. The sample was 78.3% White, 52.9% female, and
the majority of the sample (65.8%) reported having two or more years of college experience. The sample was mostly employed (92.3%), with 5.3% reporting unemployment and 2.3% reporting retirement. A very small percentage (1.8%) reported having received employment assistance from a vocational rehabilitation agency in finding their current job.

**Personality.** Personality was measured using the IPIP 20 item measures of the Big Five (Goldberg et al., 2006). The personality items can be found in the Appendix. The 20-item measures were selected over the shorter versions of the scale because MM-IRT requires a scale with a minimum of 15 items (Zickar et al., 2004). These scales measures extraversion ($\alpha = .95$ in this sample), agreeableness ($\alpha = .94$), conscientiousness ($\alpha = .93$), neuroticism ($\alpha = .96$), and openness to experience ($\alpha = .92$). Participants were asked to rate how well statements (e.g. *I am the life of the party.*) described them using a scale from $1 = \text{Strongly Disagree}$ to $5 = \text{Strongly Agree}$.

**Social Desirability.** Social desirability ($\alpha = .70$) was measured using the impression management scale developed by Paulhus (1984). This scale contains ten true or false items designed to detect socially desirable responding (e.g. *I sometimes try to get even, rather than forgive and forget*).

**Clinical Diagnoses.** Information on whether or not a participant has been diagnosed with a depressive disorder by a medical professional will be collected from the participants. Participants were asked to select whether or not they have been diagnosed by a doctor or mental health care professional with depression. This methodology has been used previously by large, longitudinal panel studies, such as the Health and Retirement Study, as a means for assessing mental health (Fisher, Faul, Weir, & Wallace, 2005). A previous depression diagnosis was reported by 25.6% of the sample.
**Depression.** Depression symptomology was measured using the CES-D ($\alpha = .94$) scale that is designed to measure depressive symptoms (Radloff, 1977). This scale consists of twenty items asking about the frequency of depressive symptoms during the last week (e.g. *During the past week I thought my life had been a failure.*) on a scale from $1 – 4$ (1 = Rarely or None of the Time (less than 1 day), 2 = Some or a Little of the Time (1-2 Days), 3 = Occasionally or a Moderate Amount of Time (3-4 days), 4 = Most or All of the Time (5-7 Days)). The items from this scale can be found in the appendix.

**Task Performance.** Because of the limitations of collecting a large sample on MTurk, task performance was measured using a self-report scale. Task performance was measured using Williams and Anderson’s (1991) measure of in-role job performance ($\alpha = .83$). This scale consists of seven items that ask about how one’s manager would rate one’s task performance (e.g. *[I] performed tasks that were expected of me.*) over the last thirty days on a scale of 1 = Strongly Disagree to 5 = Strongly Agree. These items can be found in the appendix.

**OCB.** Organizational citizenship behavior was measured using Williams and Anderson’s (1991) measure of OCB. Both the seven item individual facet (OCBI; $\alpha = .88$; e.g. *Helps others who have been absent*) and the six item organizational facet (OCBO; $\alpha = .69$; e.g. *Does not take unnecessary time off work.*) of their measure were collected. These items used the same instructions and response scale as the task performance measure.

**CWB.** Counter productive work behaviors were measured using the ten item CWB-C (Counterproductive Workplace Behaviors Checklist; $\alpha = .81$; Spector, Bauer, & Fox, 2010). Participants were asked how often they had engaged in various activities on their current job (e.g. *Came to work late without permission.*) on the following scale: 1 = Once or twice, 2 = Once or twice/month, 3 = Once or twice/week, 4 = Everyday.
Analyses

Descriptive statistics, including means, standard deviations, and reliability, were calculated for all of the scales in the study. Additionally, intercorrelations were calculated for the all scales.

CFA was used to examine both unidimensionality prior to conducting the IRT analyses (De Ayala & Hertzog, 1992) and to determine if the means could be compared between individuals with and without a previous depression diagnosis. Unidimensionality is an assumption of this IRT model, and scales that do not demonstrate sufficient unidimensional should not be examined within an IRT framework. Previous research has shown that unidimensional IRT models are fairly robust to minor violations of the assumption of unidimensionality (Drasgow, & Parsons, 1983), and that unidimensional models can be applied to multidimensional data as long as there is one primary dimension. Prior to examining mean differences on the scales, Configural, metric, and scalar invariance were examined in that order for the personality scales and the criterion scales using the R packages lavaan (Rosseel, 2012) and semTools (Pornprasertmanit et al., 2013). Fit for the CFA models was examined using several commonly used fit statistics including a chi-square significance test, Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI). Because metric and scalar invariance are nested models within the configural model they were compared to the configural model using a chi-square difference test. Configural invariance, sometimes referred to as weak factorial invariance, is used to determine if the pattern of factor loadings is consistent across groups. Without configural invariance, there is not sufficient evidence that the construct is the same across groups, this is similar to the IRT concept of differential test functioning. Metric invariance, or strong factorial invariance, tests the equivalence of factor loadings across groups.
Finally, scalar invariance tests for consistent intercepts across groups (Vandenberg & Lance, 2000). If scalar invariance is found to hold, then the means will be compared to determine if there are substantial mean differences between individuals with and without depression.

The MM-IRT-C model was analyzed using the software package Latent Gold (Vermunt & Magidson, 2005). This program was used by Tay et al. (2011) in their MM-IRT-C analyses. Once the latent classes were extracted using the MM-IRT-C model, the classes that emerged were correlated with the criterion measures, including task performance, OCBI, OCBO and CWB to determine if the relationship between performance and personality was consistent across the latent classes.
CHAPTER 4. RESULTS

Descriptive Analyses

Means, standard deviations, and reliabilities for the scales can be found in Table 1. Table 1 also contains the correlations among all of scales included in the study. Because twenty-five percent of the sample reported having been diagnosed with depression by a doctor, the diagnostic measure of depression is the focus of the analyses for this paper rather than scores on the CESD. A diagnosis of depression is more definitive than a self-report measure of symptomology and is less likely to be subject to common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). A previous depression diagnosis and CESD scores were positively related to one another ($r = .36, p < .01$), and showed the same pattern of relationships with the other variables in the study. CESD tended to be more strongly correlated with all other study variables than a depression diagnosis, supporting the possibility that CESD was subject to common method bias.

The results in Table 1 provide support for several hypotheses. Hypothesis 2 predicted that openness would be uncorrelated with depression, and as expected depression was uncorrelated with Openness to experience ($r = .01, p > .05$). Hypothesis 5 hypothesized that depression was positively correlated with neuroticism, which was also confirmed ($r = .35, p < .01$). Depression was negatively correlated with conscientiousness ($r = -.17, p < .01$) and extraversion ($r = -.18, p < .01$) confirming hypothesis 7, which hypothesized a negative correlation between depression and conscientiousness, and hypothesis 9, which hypothesized a negative relationship between depression and extraversion. Agreeableness was negatively correlated with depression ($r = -.06, p < .05$), which failed to support Hypothesis 4; however it should be noted that the correlation between agreeableness and a depression diagnosis was quite small and explained less than one percent of the variance in agreeableness ($R^2 < .01$). As expected depression was negatively
correlated with task performance, which supported hypothesis 13, which predicted they would be negatively correlated; however the size of the correlation was fairly small \( r = -.07, p < .05 \)

**Invariance Analyses and Mean Differences**

The next step was to look at the measurement invariance in personality across participants with and without depression. The model included all five personality constructs, conscientiousness, extraversion, emotional stability, agreeableness, and openness. First, each group was examined independently in a CFA framework in order to determine how well the model fit each group. The model showed mixed evidence of acceptable fit for the participants with depression. The chi-square was significant \( \chi^2 (4840) = 10549.70, p < .01 \), but the RMSEA was acceptable (RMSEA=.07). The CFI was also not acceptable (CFI=.62). In an effort to improve the fit, modification indices were examined. Modifications were made one at a time until the addition of a new constraint failed to improve the fit. Specifically, several items within scales showed higher correlations than would be expected solely due to the presence of a latent trait suggesting local dependence. Allowing the errors of these items to correlate with one another helps to improve model fit. The final model allowed the following errors to be correlated: Conscientiousness10 with Conscientiousness4, Emotional_ Stability10_R_r with Emotional_ Stability11_R_r, Openness16_R_r with Openness2, Emotional_ Stability13_R_r with Emotional_ Stability2, Openness12 with Openness19_R_r, Conscientiousness12_R_r with Conscientiousness20_R_r, and Openness1 with Openness5. The new model still showed mixed evidence of fit. The chi-square was significant \( \chi^2 (4950) = 19873.20, p < .01 \), but the RMSEA was acceptable (RMSEA=.07). The CFI was also not acceptable (CFI=.67).
Turning to the model examining the participants who did not report a diagnosis of depression, mixed evidence was found for acceptable fit. The chi-square was significant ($\chi^2 (4840) = 17353.26, p < .01$), but the RMSEA was acceptable (RMSEA=.06). The CFI was not acceptable for this model (CFI=.72). This model was also run a second time with the same correlated errors as were present in the model examining participants with depression. This model again showed an improvement over the original model but still showed mixed evidence of fit. The chi-square was significant ($\chi^2 (4833) = 15913.08, p < .01$), but the RMSEA was acceptable (RMSEA=.06). The CFI was also not acceptable (CFI=.75). Because both the depressed and nondepressed models showed at least some evidence of acceptable fit, I moved on to examining configural invariance using the second model that contained the additional correlated errors.

The configural invariance model also showed mixed evidence of acceptable model fit. The chi-square was significant ($\chi^2 (9666) = 25744.00, p < .01$), but the RMSEA was acceptable (RMSEA=.06). The CFI was not acceptable for this model (CFI=.73). Given the mixed evidence for acceptable model fit, it is unclear if the factor structure is equivalent between participants with and without depression. Because the evidence was mixed, I decided to move forward with testing the next level of invariance, metric invariance.

The metric invariance model was not supported. The model again showed mixed evidence of acceptable fit. The chi-square was significant ($\chi^2 (9761) = 25920, p < .01$), but the RMSEA was acceptable (RMSEA=.06). The CFI was not acceptable for this model (CFI=.73). The chi-square difference test was significant indicating that the metric invariance model was significantly worse than the configural model ($\Delta \chi^2 (95) = 175.71, p < .01$). Because the metric invariance model was not supported, I did not test scalar invariance, because this is a stricter
model than metric invariance. Because scalar invariance could not be shown, it is not reasonable to compare the means between participants with and without depression.

The criterion measures were also checked for invariance to ensure that task performance, CWB, and OCB were equivalent between the two groups. First, task performance was checked for invariance. The model fit for participants with depression showed mixed evidence of acceptable model fit. The chi-square was significant ($\chi^2 (14) = 62.47, p < .01$), and the RMSEA was not acceptable (RMSEA=.12). However, the CFI was acceptable for this model (CFI=.95). The model fit for non-depressed participants also showed mixed evidence of acceptable fit. The chi-square was significant ($\chi^2 (14) = 177.54, p < .01$), and the RMSEA was not acceptable (RMSEA=.13). However, the CFI was acceptable for this model (CFI=.95).

Even though the model fit was mixed for the individual groups, I moved on to testing configural invariance. The chi-square was significant ($\chi^2 (28) = 240.01, p < .01$), and the RMSEA was not acceptable (RMSEA=.13). However, the CFI was acceptable for this model (CFI=.95). The chi-square difference test was significant indicating that the metric invariance model was significantly worse than the configural model ($\Delta \chi^2 (95) = 179.71, p < .01$). Because the metric invariance model was not supported for task performance, I did not test scalar invariance.

For CWB the model fit the data for the depressed participants slightly better than the non-depressed participants. With the depressed participants, the chi-square was significant ($\chi^2 (35) = 281.38, p < .01$), the RMSEA was acceptable (RMSEA=.08), and the CFI was close to acceptable (CFI=.92). For the non-depressed participants the chi-square was significant ($\chi^2 (35) = 279.63, p < .01$), the RMSEA was not acceptable (RMSEA=.10), and the CFI was not
acceptable (CFI=.87). The configural invariance analysis showed mixed support for configural invariance. The chi-square was significant ($\chi^2 (70) = 361.00, p < .01$), the RMSEA was not acceptable (RMSEA=.10), and the CFI was not acceptable (CFI=.88). The chi-square difference test for metric invariance showed a lack of support for metric invariance ($\Delta \chi^2 (9) = 56.81, p < .01$).

Finally, OCBI and OCBO were examined in a CFA model. For the depressed participants the chi-square was significant ($\chi^2 (64) = 225.28, p < .01$), the RMSEA was not acceptable (RMSEA=.10), and the CFI was close to acceptable (CFI=.85). For the non-depressed participants the chi-square was significant ($\chi^2 (64) = 426.63, p < .01$), the RMSEA was not acceptable (RMSEA=.09), and the CFI was close to acceptable (CFI=.90). The configural invariance analysis showed mixed support for configural invariance. The chi-square was significant ($\chi^2 (128) = 651.91, p < .01$), the RMSEA was not acceptable (RMSEA=.09), and the CFI was not acceptable (CFI=.89). The chi-square difference test for metric invariance showed support for metric invariance ($\Delta \chi^2 (11) = 7.69, p = .74$). Because metric invariance was supported I moved on to testing scalar invariance. Scalar invariance was also supported ($\Delta \chi^2 (11) = 18.55, p = .07$) indicating that the OCB scales were invariant between those with and without depression.

**MM-IRT-C**

Next, MM-IRT-C analyses were run for each of the personality scales. First, a series of models with an increasing number of latent classes were run to determine how many latent classes were present in each construct. The BIC statistic was examined for each additional model in accordance with previous research (Carter et al., 2011; Tay & Vermunt, 2011) and the final
model was selected when the addition of a latent class resulted in worse fit than the previous model. Once the number of latent classes had been established, a new model was run containing the final number of latent classes and inserting social desirability and depression as drivers of the latent classes. If the new model showed better fit, then the regression weights and significance of social desirability and depression were examined in more detail, if the new model showed poorer fit then it was determined that the more parsimonious model was the final model. Fit statistics for all of the models discussed below can be found in Table 2.

Openness was analyzed with one (O1; $BIC = 17658.29$), two (O2; $BIC = 17612.30$), and three (O3; 17625.03) classes, with the two class model showing the best fit out of the three. Social desirability (O4, $BIC = 17607.94$) and depression (O5, $BIC = 17613.87$) were then modeled in turn as covariates in the two-class model. The final model containing both depression and social desirability could not be estimated due to an error in Latent Gold that caused the program to crash whenever this model was attempted. Based on the fit statistics from the depression only and the social desirability only models, model O4 was selected as the final model. This provides support for Hypothesis 1 because depression was not a significant covariate in this model, as evidenced by the lack of improved fit for model O5. The latent classes in the final model (Figure 1) are almost completely overlapping for all items. Social desirability had a positive regression coefficient for class 1 ($\beta = 1.86$, $z(2.03)$, $p = .04$) and a negative coefficient for class 2 ($\beta = -1.78$, $z(-6.44)$, $p < .01$). This opposite effect of social desirability according to group membership has the effect of pushing the two classes together.

Like the previous construct, agreeableness was tested against a one (A1; $BIC = 16974.00$), two (A2; $BIC = 16969.31$), and three (A3; $BIC = 16976.81$) class model, and the two class model was shown to have the best fit. The two class model was then tested with social desirability (A5;
BIC = 16954.22) and depression (A6; BIC = 17104.17) as unique covariates, then with both covariates in the same model (A4; BIC = 16962.69). Model A5 (BIC = 16954.22) in Table 2 contains the final agreeableness model, which has two latent classes with social desirability as a significant covariate. This provides support for Hypothesis 3 because depression was not a significant covariate in the model. Social desirability had positive regression coefficients for both latent class 1 ($\beta = 3.50$, $z(8.96)$, $p < .01$) and latent class 2 ($\beta = 1.11$, $z(5.10)$, $p < .01$) with class 1 having the stronger coefficient. Figure two contains the mean response patterns for the two groups. Class 1 has a lower mean across all items; however, the stronger social desirability scores have pushed the less agreeable group up towards the more agreeable class 2.

Like the previous two constructs, neuroticism was fit with a one (N1; BIC = 17624.71), two (N2; BIC = 17566.87), and three (N3; BIC = 17572.44) class model with the two class model showing the best fit. The two class model was then fit with both covariates (N4; BIC = 17571.14), social desirability only (N5; BIC = 17641.02), and depression only (N6; BIC = 17704.00); however, none of these models showed an increase in fit over the more parsimonious two class model indicating that the covariates were not significant for this model. The response patterns for the final model (N2) can be found in Figure 3. Class 1 is characterized by lower means on most items, whereas class 2 tended to show higher means with the exception of the item [I] Worry about things. These results failed to support hypothesis 6 that depression would emerge as an explanation for the latent classes in neuroticism.

Conscientiousness was fit with a one (C1; BIC = 17672.05), two (C2; BIC = 17640.04), and three (C3; BIC = 18607.20) class model with the two class model showing the best fit. The two class model was the run with both covariates (C4; BIC = 17636.63), social desirability only (C5; BIC = 17661.91), and depression only (C6; BIC = 17689.12). The model with social
desirability and depression as covariates showed the best fit. This provides support for hypothesis 8 that depression would emerge as a covariate in the conscientiousness model. Social desirability negatively predicted membership in both class 1 ($\beta = -1.05$, $z(-2.75)$, $p < .01$) and class 2($\beta = -1.59$, $z(-6.93)$, $p < .01$). Depression had a negative weight for class 1($\beta = -6.44$, $z(-5.95)$, $p < .01$), indicating that an absence of depression made one more likely to fall into class 1, and was non-significant for class 2 ($\beta = .69$, $z(1.95)$, $p = .05$). Figure 4 contains the mean response patterns for the latent classes. Class two showed lower means scores than class 1.

Extraversion was fit with one (E1; $BIC = 18785.78$), two (E2; $BIC = 18690.36$), three (E3; $BIC = 18680.00$), and four (E4; $BIC = 18693.47$) class models with the three class model showing the best fit. The three class model was then tested with social desirability (E5; $BIC = 18633.73$) and depression (E6; $BIC = 18692.12$) as unique covariates. The combined covariate model could not be computed due to a Latent Gold error. The three class model with social desirability as a covariate (E5) was selected as the final model. This failed to find support for hypothesis 10; depression was not a significant covariate. Social desirability was significant predictor for class 1 ($\beta = -.96$, $z(-3.98)$, $p < .01$), class 2($\beta = 1.65$, $z(6.52)$, $p < .01$), and class 3 ($\beta = 2.25$, $z(7.85)$, $p < .01$). Figure 5 contains the mean response patterns for the three classes. When the classes differed from one another, class three tended to be the highest scorers, followed by class 1, then class 2.

Looking across the personality constructs, these results provide partial support for hypothesis that social desirable responding would contribute to the latent classes. Four out of the five constructs, openness, agreeableness, conscientiousness, and extraversion, included social desirability as a covariate in the final model. Neuroticism was the only construct that did not have a class related to socially desirable responding.
The predicted classes for each participant were saved from the five final MM-IRT-C models to examine if class membership was related to task performance, OCBI, OCBO, and CWB. Because conscientiousness was the only construct that contained latent classes that were related to depression, the discussion of hypotheses 12a – 12c focuses on conscientiousness. However, as an exploratory question the relationship between the latent classes and performance will be explored for all constructs. The relationships among the latent classes and the criterion variables can be found in Table 3.

Task performance was negatively correlated with conscientiousness ($r = -.17, p < .01$) classification such that individuals in class 2, the class characterized by lower conscientiousness scores, also showed lower performance. Class 1, which was also characterized by a lack of depression, showed both higher conscientiousness scores and higher task performance scores. This provides support for hypothesis 12a in that a lack of depression predicted higher task performance. OCBI ($r = -.15, p < .01$) and OCBO ($r = -.25, p < .01$) showed the same pattern in that those in class 1 had higher scores on both measure of OCB providing support for hypothesis 12b. Conscientiousness class member was positively correlated with CWB ($r = .11, p < .01$), such that those class 1 had lower scores on CWB, supporting hypothesis 12c.

Neuroticism was correlated with OCBI ($r = .08, p < .01$) and OCBO ($r = .07, p < .05$), such that those in class two, the more emotionally stable class, showed higher scores on both OCBI and OCBO. Agreeableness class membership was also correlated with OCBI ($r = .10, p < .01$) such that class two, the class characterized by very high social desirability scores, also scored higher on the measure of OCBI. Extraversion class membership was negatively correlated with CWB ($r = -.10, p < .01$) such that those in class 3, the class with the highest social
desirability scores, reported the least amount of counterproductive behavior. Openness class membership was unrelated to all of the criterion measures.
CHAPTER 5. DISCUSSION

This paper has examined the relationship between personality and depression, specifically if individuals with a prior depression diagnosis respond to personality scales differently than their non-depressed counterparts and if this could lead to unintended discrimination when personality scales are used in a selection context. Because depression is a mental illness, it is covered under the American’s with Disabilities Act and cannot be used to discriminate in hiring decisions. By better understanding the relationship among depression diagnoses, personality, and job performance, it is possible to determine if accidental discrimination is occurring, and if so, to prevent discrimination by building scales that are less biased but that still predict performance.

There has been previous research regarding the relationship between personality scales and mental illnesses, including depression. The correlations between the personality scales and depression that were examined largely confirmed previous research. Depression was positively correlated with neuroticism, uncorrelated with openness, and was negatively correlated with conscientiousness and extraversion in accordance with previous research and meta-analyses (Anderson & McLean, 1997; Hirschfeld, et al., 1983; Jylhä, & Isometsä, 2006; Kotov et al., 2010; Liebowitz et al., 1979; Malouff et al., 2005). Depression was also negatively correlated with task performance in support of previous research (Adler et al., 2006; Lerner & Henke, 2008; Wang et al., 2004). The only correlation that failed to support the existing literature was agreeableness, which was negatively correlated with depression instead of uncorrelated as has been previously found in two meta-analyses (Kotov et al., 2010; Malouff et al., 2005). This could be a type I error in that an effect was incorrectly found that did not exist, or it could be due to the sample size in the current research. The current sample used a large sample size, which can cause
even small correlations to become statistically significant. The variance in agreeableness that was explained by a previous depression diagnosis was less than one percent ($R^2 < .01$), which indicates that even though the relationship was statistically significant it may not be practically significant.

Once the pattern of correlations had been established, confirmatory factor analyses (CFA) were examined for all of the personality constructs and the criterion constructs as well. The goal of running a CFA for the personality constructs was to determine if means could reasonably be compared between those with and without depression, and to determine if the constructs displayed sufficient unidimensionality to justify using a unidimensional IRT model. Measurement invariance analyses were conducted on the criterion measures to understand if there were underlying differences between how individuals with and without a prior depression diagnosis were responding to the CWB, OCB, and task performance measures. Differences in how the criterion scales are used by the two groups could lead to inaccurate conclusions if, for example, there is not a true difference in task performance between the groups, but merely a difference in how depressed individuals and non-depressed individuals use the task performance scale.

The personality scales did not show metric invariance and accordingly mean differences were not analyzed between participants with and without a prior depression diagnosis. This indicates that these scales were performing differently between the groups and provides additional justification for the MM-IRT-C analyses that were conducted. The fit did improve slightly when the errors of a few items were allowed to correlate; however the improvement in overall fit was not sufficient to make the model fit well or to support a higher level of invariance. The lack of invariance indicates that these five personality scales as a whole function differently
between the groups with and without a previous depression diagnosis, and prevented the examination of mean differences between the two groups. The measurement invariance analyses do not conclusively show that a depression diagnosis is the cause of the lack of invariance, only that there are differences in how the scales are used between the two groups. The lack of invariance calls into question any mean differences that have been previously found between individuals with and without a depression diagnosis using the IPIP scales. If the two groups are using the scales differently then it impossible to tease apart true mean differences from differences that are due to the lack of invariance in the scales.

Configural invariance was found to hold for task performance and CWB, and scalar invariance was found to hold for OCBI and OCBO. This indicates that the task performance and CWB related results comparing the two groups on conscientiousness should be taken with a grain of salt because the two groups were using the task performance and CWB scales in different ways. However, it is encouraging that they show a similar pattern of results to the invariant OCBI and OCBO measures.

Turning to the MM-IRT-C analyses that were conducted, mixed evidence was found for the hypothesis that a previous depression diagnosis would drive latent classes for the constructs where depression had previously been shown to impact average scores, but most of the analyses supported the hypothesis that socially desirable responding would drive latent class membership. In the openness scale, higher scores on social desirability characterized class 1 whereas class 2 membership was predicted by a lack of social desirability (i.e. honest responding).

Agreeableness also showed latent classes that were predicted by social desirability; however, both classes were characterized by high levels of social desirability. In this case class 1 was the highly socially desirable class compared to class 2, which was the somewhat socially
desirable class. Extraversion showed a similar pattern of results with class 1 as the honest class, class 2 as the moderately socially desirable class, and class 3 as the highly socially desirable class.

Neuroticism showed two latent classes; however, contrary to what was hypothesized, these classes could not be explained by either social desirability or depression. It is possible that depression was not driving latent classes because neuroticism scores are not dependent on the presence of depression. Although depression and neuroticism are correlated, the presence of depression did not change the way participants responded to the items on the neuroticism scale. Previous research examining latent classes in the construct of neuroticism has also found unexpected and incongruous results. Maij-de Meij and colleagues (2008) found that neuroticism contained three latent classes, and unlike the present research, found that social desirability was a significant predictor of the classes that emerged. Zickar and colleagues (2004) found that neuroticism was the only personality construct to contain four latent classes. It is unclear why these two papers and the present research found such different results for the number of latent classes and the significance of social desirability in neuroticism when the findings of the present research were consistent with past research for the other four personality constructs. All three papers used different batteries to assess personality, but this seems an unlikely explanation for the divergent findings given that the other four sets of findings were consistent. Future research should continue to explore the number of latent classes in neuroticism, whether or not social desirability impacts latent class membership, and whether or not depression impacts latent class membership.

The conscientiousness model showed two latent classes that were explained by both social desirability and depression. Class 1 was the somewhat honest non-depressed class;
whereas class 2 was the extremely honest, more likely to be depressed class. The conscientiousness classes predicted the criterion data as expected. Class 2 membership was negatively related to task performance, OCBO, and OCBI, and class 2 membership was also positively related to CWB. This provides evidence that a previous depression diagnosis, in conjunction with social desirability, is driving some of the latent class membership for conscientiousness. Furthermore, those in the non-depressed class had higher scores on the criterion measures. This suggests that the presence of a previous depression diagnosis is changing how the scales are used, and that the class that is more likely to have been diagnosed with depression will also be worse performers in terms of task performance, OCBO, and OCBI in addition to showing higher levels of CWB.

It is unclear why conscientiousness was the only construct to show latent classes that could be explained by a previous depression diagnosis. Conscientiousness, neuroticism, and extraversion were the constructs that were hypothesized to show latent classes due to depression. Bienvenu and colleagues’ (2004) research on subfacet differences due to mental illness may help provide an explanation for the pattern of results that emerged. In their research neuroticism show significant mean differences across all subfacets. The size of the mean differences changed across subfacets. I expected that this would lead to latent classes that could be explained by depression; however it appears that the significant differences across all six subfacets were too consistent to explain latent class membership as expected. For extraversion and conscientiousness, they found only one of the subfacets, self-discipline for conscientiousness and assertiveness for extraversion, to display significant differences between those with and without depression. This would suggest that both extraversion and conscientiousness should have displayed latent classes that could be explained by a depression diagnosis to the extent that the
scales contained items tied to assertiveness and self-discipline, but a depression diagnosis only explained the latent classes that emerged in conscientiousness.

**Practical Implications**

These results have implications for the use of personality by practitioners in selection decisions. First, ever having been diagnosed with depression was unrelated to latent class membership for agreeableness, emotional stability, openness, and extraversion suggesting that these tests do not introduce bias due to response patterns. However, the overall CFA did show a lack of invariance across all five scales, which suggests that those with and without a prior depression diagnosis may be using the scales differently. It should be noted that this may not be due to the presence of a previous depression diagnosis. It could be due to another unmeasured variable that was not accounted for in the CFA, or it could also be due to the large sample size in the group without a previous depression diagnosis. This mixed set of results suggests that a depression diagnosis does not account for the how these scales are used above and beyond the explanation provided by latent class membership; however, the full personality battery does not function equivalently between the two groups. The lack of overall invariance could be due in part to the latent class differences found in conscientiousness. Overall, a previous depression diagnosis does not seem to be a strong driver of personality scale usage, and should not be a large concern when designing and selecting personality tests with the exception of conscientiousness, which should be explored further.

These results suggest that personality scales as a whole may not be the biggest source of concern for whether or not individuals with depression are able to obtain jobs, although the IPIP conscientiousness scale did contain latent classes that could be explained by the presence of previous depression diagnosis. Depression is more likely to impact employment at other stages in
the job search and hiring process. For example, there has been previous research that has shown that depression is negatively related to gaining new employment after a job loss (Lerner & Henke, 2008). Individuals with depression that is severe enough to prevent them from beginning the job application process will not make it to the stage where they are completing selection tests, such as personality tests.

Depression could also be problematic during the interview stage of the hiring process if it results in an applicant appearing uninterested in the job opening or if the applicant is unable to attend the interview due to a depressive episode. Finally, depression has been linked to absenteeism, presenteeism, and poor job performance (Ettner, Maclean, & French, 2011; Lerner & Henke, 2008; Lim, Sanderson, & Andrews, 2000). A job applicant with depression may be less likely to receive a positive reference due to performance on a previous job, or may be more likely to have been let go from a previous job due to performance or attendance issues. In summary, practitioners should not view personality testing a large source of concern when trying to design tests and selection systems that are fair to individuals with and without a previous depression diagnosis.

**Limitations and Future Directions**

This paper contains several limitations that should be addressed in future research. First, due to the constraints of the sample that was collected, it was not possible to use supervisor ratings of performance or receive confirmation of the presence or absence of depression in each participant from a licensed physician. Supervisor ratings would ensure that the task performance data is as accurate as possible. Although supervisors can still fail to provide perfectly accurate performance data, they do tend to provide more accurate data than the supervisees. Additionally, collecting depression data from a physician would prevent participants from self-diagnosing or
from saying they had never been diagnosed because they did not feel comfortable disclosing. The participants did have the option to skip questions that they did not wish to answer, although this option was used by only two percent of the sample when responding to the depression diagnosis item, but some participants who did not wish to disclose their depression status could have still responded in a socially desirable manner instead of skipping the depression diagnosis question. Physician data could also have allowed for a sample that was known to be currently suffering from depression instead of looking at lifetime diagnoses.

Another limitation of this research was using a lifetime measure of depression instead of a momentary measure of depression. The depression diagnosis item that was used asked about lifetime diagnoses instead of whether or not the participant was currently experiencing a depressive episode. This approach was taken for two reasons; first of all, there has been research to show that depression can impact personality even when someone is not actively suffering from a depressive episode (e.g. Anderson & McLean, 1997). Secondly, individuals with chronic depression may not return to a doctor for confirmation each and every time they have a depressive episode. An item that looked at diagnoses within the last year could have incorrectly excluded participants who had been diagnosed over a year ago, but who were still actively suffering from depression.

The sample also presented some demographic limitations that could prevent this research from being generalized to the population of the United States. Specifically, the sample was more educated than the United States population. According to the United States Census Bureau (2015) 58.9% of the US population over the age of 25 had completed one or more years of college in 2015. This is a much lower level of educational attainment than was reported by the sample in the present research (92.3%). Additionally, 92.3% of the present sample reported
being employed compared to 63% of the population that was reported as being in the US labor force in 2016 (Bureau of Labor Statistics, 2017).

Finally, the present research is limited because they used one inventory for each of the big five instead of looking at a variety of different scales for each construct. These results may not generalize to other publicly available personality scales or to proprietary scales that utilize different items. Examining a wide variety of scales that measure the Big Five is beyond the scope of this research; however it should be noted that future research should examine whether or not a depression diagnosis can explain latent classes that emerge on personality inventories other than the IPIP.

Future research should look at how classification on the measure of conscientiousness changes within person longitudinally before, during, and after a depressive episode. There has been research looking at mean changes in conscientiousness scores (Anderson & McLean, 1997); however, it would be helpful to know if an individual simply shifted his or her overall mean or if his or her response pattern actually changed during a depressive episode. Future research should also examine latent classes at the subfacet level for conscientiousness. It is possible that some subfacets of conscientiousness, such as self-discipline, which has been previously shown to have mean differences on depression (Bienvenu et al., 2004), are more impacted by these classes than others. This will be important to understand for practitioners who measure conscientiousness at the subfacet level. Finally, future research should look into other conscientiousness scales and consider testing revised conscientiousness scales that remove items that were shown to have greater mean differences and should calculate adverse impact ratios in a selection scenario to determine how much adverse impact conscientiousness is introducing into selection systems in practice.
REFERENCES


APPENDIX A. INSTRUMENT

Personality Items
Agreeableness1 Am interested in people.
Agreeableness10 Am on good terms with nearly everyone.
Agreeableness11 Have a good word for everyone.
Agreeableness12 Show my gratitude.
Agreeableness13 Think of others first.
Agreeableness14 Love to help others.
Agreeableness15_R Insult people.
Agreeableness16_R Am not interested in other people's problems.
Agreeableness17_R Feel little concern for others.
Agreeableness18_R Am not really interested in others.
Agreeableness19_R Am hard to get to know.
Agreeableness2 Sympathize with others' feelings.
Agreeableness20_R Am indifferent to the feelings of others.
Agreeableness3 Have a soft heart.
Agreeableness4 Take time out for others.
Agreeableness5 Feel others' emotions.
Agreeableness6 Make people feel at ease.
Agreeableness7 Inquire about others' well-being.
Agreeableness8 Know how to comfort others.
Agreeableness9 Love children.
Conscientiousness1 Am always prepared.
Conscientiousness10 Love order and regularity.
Conscientiousness11 Like to tidy up.
Conscientiousness12_R Leave my belongings around.
Conscientiousness13_R Make a mess of things.
Conscientiousness14_R Often forget to put things back in their proper place.
Conscientiousness15_R Shirk my duties.
Conscientiousness16_R Neglect my duties.
Conscientiousness17_R Waste my time.
Conscientiousness18_R Do things in a half-way manner.
Conscientiousness19_R Find it difficult to get down to work.
Conscientiousness2 Pay attention to details.
Conscientiousness20_R Leave a mess in my room.
Conscientiousness3 Get chores done right away.
Conscientiousness4 Like order.
Conscientiousness5 Follow a schedule.
Conscientiousness6 Am exacting in my work.
Conscientiousness7 Do things according to a plan.
Conscientiousness8 Continue until everything is perfect.
Conscientiousness9 Make plans and stick to them.
Emotional Stability1 Am relaxed most of the time.
Emotional Stability10_R Change my mood a lot.
Emotional Stability11_R Have frequent mood swings.
Emotional Stability12_R Get irritated easily.
Emotional Stability13_R Often feel blue.
Emotional Stability14_R Get angry easily.
Emotional Stability15_R Panic easily.
Emotional Stability16_R Feel threatened easily.
Emotional Stability17_R Get overwhelmed by emotions.
Emotional Stability18_R Take offense easily.
Emotional Stability19_R Get caught up in my problems.
Emotional Stability2 Seldom feel blue.
Emotional Stability20_R Grumble about things.
Emotional Stability3 Am not easily bothered by things.
Emotional Stability4 Rarely get irritated.
Emotional Stability5 Seldom get mad.
Emotional Stability6_R Get stressed out easily.
Emotional Stability7_R Worry about things.
Emotional Stability8_R Am easily disturbed.
Emotional Stability9_R Get upset easily.
Extraversion1 Am the life of the party.
Extraversion10 Am skilled in handling social situations.
Extraversion11_R Don't talk a lot.
Extraversion12_R Keep in the background.
Extraversion13_R Have little to say.
Extraversion14_R Don't like to draw attention to myself.
Extraversion15_R Am quiet around strangers.
Extraversion16_R Find it difficult to approach others.
Extraversion17_R Often feel uncomfortable around others.
Extraversion18_R Bottle up my feelings.
Extraversion19_R Am a very private person.
Extraversion2 Feel comfortable around people.
Extraversion20_R Wait for others to lead the way.
Extraversion3 Start conversations.
Extraversion4 Talk to a lot of different people at parties.
Extraversion5 Don't mind being the center of attention.
Extraversion6 Make friends easily.
Extraversion7 Take charge.
Extraversion8 Know how to captivate people.
Extraversion9 Feel at ease with people.
Openness1 Have a rich vocabulary.
Openness 10  Can handle a lot of information.
Openness 11  Love to think up new ways of doing things.
Openness 12  Love to read challenging material.
Openness 13  Am good at many things.
Openness 14  Have difficulty understanding abstract ideas.
Openness 15  Am not interested in abstract ideas.
Openness 16  Do not have a good imagination.
Openness 17  Try to avoid complex people.
Openness 18  Have difficulty imagining things.
Openness 19  Avoid difficult reading material.
Openness 2  Have a vivid imagination.
Openness 20  Will not probe deeply into a subject.
Openness 3  Have excellent ideas.
Openness 4  Am quick to understand things.
Openness 5  Use difficult words.
Openness 6  Spend time reflecting on things.
Openness 7  Am full of ideas.
Openness 8  Carry the conversation to a higher level.
Openness 9  Catch on to things quickly.

Depression
CESD1  I was bothered by things that usually don't bother me.
CESD2  I did not feel like eating; my appetite was poor.
CESD3  I felt that I could not shake off the blues even with help from my family or friends.
CESD4  I felt that I was just as good as other people.
CESD5  I had trouble keeping my mind on what I was doing.
CESD6  I felt depressed.
CESD7  I felt that everything I did was an effort.
CESD8  I felt hopeful about the future.
CESD9  I thought my life had been a failure.
CESD10  I felt fearful.
CESD11  My sleep was restless.
CESD12  I was happy.
CESD13  I talked less than usual.
CESD14  I felt lonely.
CESD15  People were unfriendly.
CESD16  I enjoyed life.
CESD17  I had crying spells.
CESD18  I felt sad.
CESD19  I felt that people dislike me.
CESD20  I could not get "going."
Social Desirability

Social_Desirability1_R It is sometimes hard for me to go on with my work if I am not encouraged.
Social_Desirability2_R I sometimes feel resentful when I don't get my way.
Social_Desirability3 No matter who I'm talking to, I'm always a good listener.
Social_Desirability4 There have been occasions when I took advantage of someone.
Social_Desirability5 I'm always willing to admit it when I make a mistake.
Social_Desirability6_R I sometimes try to get even rather than forgive and forget.
Social_Desirability7 I am always courteous, even to people who are disagreeable.
Social_Desirability8 I have never been irked when people expressed ideas very different from my own.
Social_Desirability9_R There have been times when I was quite jealous of the good fortune of others.
Social_Desirability10_R I am sometimes irritated by people who ask favors of me.
Social_Desirability11 I have never deliberately said something that hurt someone's feelings.

Criterions Items (Task Performance, CWB, OCB)

Task1 adequately completed my assigned duties.
Task2 fulfilled all responsibilities specified in my job description.
Task3 performed tasks that were expected of me.
Task4 met formal performance requirements of the job.
Task5 engaged in activities that directly affected my performance evaluation.
Task6_R neglected aspects of the job I am obligated to perform.
Task7_R failed to perform essential duties
OCBI_1 Helps others who have heavy work loads
OCBI_2 Helps others who have been absent
OCBI_3 Assists supervisor with his/her work (when not asked)
OCBI_4 Takes time to listen to co-workers' problems and worries
OCBI_5 Goes out of way to help new employees
OCBI_6 Takes a personal interest in other employees
OCBI_7 Passes along information to co-workers
OCBO_1 Adheres to informal rules devised to maintain order
OCBO_2_R Complains about insignificant things at work
OCBO_3_R Great deal of time spent with personal phone conversations
OCBO_4_R Takes undeserved work breaks
OCBO_5 Gives advance notice when unable to come to work
OCBO_6 Attendance at work is above the norm.
CWB1 Purposely wasted your employer’s materials/supplies
CWB2 Complained about insignificant things at work
CWB3 Told people outside the job what a lousy place you work for
CWB4 Came to work late without permission
CWB5 Stayed home from work and said you were sick when you weren’t
CWB6    Insulted someone about their job performance
CWB7    Made fun of someone’s personal life
CWB8    Ignored someone at work
CWB9    Started an argument with someone at work
CWB10   Insulted or made fun of someone at work

Attention Checks
Attn1    Occasionally eat cement.
Attn2    Select strongly disagree.
Attn3    Choose the middle response.
Attn4    am a toy poodle.
Attn5    Leave this item blank.
Attn6    Select everyday.
Attn7    Select rarely or none of the time (Less than 1 day)
DATE: June 22, 2016

TO: Rachel King, MA
FROM: Bowling Green State University Human Subjects Review Board

PROJECT TITLE: [905322-1] Using MM-IRT-C to Explore the Relationship between Depression and Pre-Employment Personality Tests

SUBMISSION TYPE: New Project

ACTION: DETERMINATION OF EXEMPT STATUS
DECISION DATE: June 20, 2016

REVIEW CATEGORY: Exemption category #2

Thank you for your submission of New Project materials for this project. The Bowling Green State University Human Subjects Review Board has determined this project is exempt from IRB review according to federal regulations AND that the proposed research has met the principles outlined in the Belmont Report. You may now begin the research activities.

Reviewer Comment: Last line of the first questionnaire - occasionally is spelled incorrectly.

Note that an amendment may not be made to exempt research because of the possibility that proposed changes may change the research in such a way that it is no longer meets the criteria for exemption. A new application must be submitted and reviewed prior to modifying the research activity, unless the researcher believes that the change must be made to prevent harm to participants. In these cases, the Office of Research Compliance must be notified as soon as practicable.

We will retain a copy of this correspondence within our records.

If you have any questions, please contact Kristin Hagemyer at 419-372-7716 or khagemy@bgsu.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within Bowling Green State University Human Subjects Review Board’s records.
Table 1. Means, Standard Deviations, Reliabilities, and Inter-scale Correlations (n = 947 - 1026)

<table>
<thead>
<tr>
<th>Factor</th>
<th>M</th>
<th>SD</th>
<th>α</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>3.87</td>
<td>.67</td>
<td>.93</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Neuroticism</td>
<td>2.43</td>
<td>.89</td>
<td>.96</td>
<td>-.48**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Extraversion</td>
<td>3.04</td>
<td>.87</td>
<td>.95</td>
<td>.33**</td>
<td>-.51**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Openness</td>
<td>3.89</td>
<td>.64</td>
<td>.92</td>
<td>.33**</td>
<td>-.37**</td>
<td>.42**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Agreeableness</td>
<td>3.83</td>
<td>.70</td>
<td>.94</td>
<td>.42**</td>
<td>-.39**</td>
<td>.49**</td>
<td>.43**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. CESD</td>
<td>1.61</td>
<td>.57</td>
<td>.94</td>
<td>-.45**</td>
<td>.66**</td>
<td>-.43**</td>
<td>-.24**</td>
<td>-.32**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Depression diagnosis</td>
<td>.26</td>
<td>.44</td>
<td>-</td>
<td>-.17**</td>
<td>.35**</td>
<td>-.18**</td>
<td>.01</td>
<td>-.06*</td>
<td>.36**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Social Desirability</td>
<td>.49</td>
<td>.24</td>
<td>.70</td>
<td>.30**</td>
<td>-.43**</td>
<td>-.27**</td>
<td>.17**</td>
<td>.39**</td>
<td>-.29**</td>
<td>-.17**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. CWB</td>
<td>1.20</td>
<td>.31</td>
<td>.81</td>
<td>-.33**</td>
<td>.34**</td>
<td>-.16**</td>
<td>-.21**</td>
<td>-.32**</td>
<td>.39**</td>
<td>.11**</td>
<td>-.22**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. OCB-I</td>
<td>5.42</td>
<td>.99</td>
<td>.88</td>
<td>.38**</td>
<td>-.29**</td>
<td>.39**</td>
<td>.44**</td>
<td>.63**</td>
<td>-.24**</td>
<td>-.02</td>
<td>.27**</td>
<td>-.25**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. OCB-O</td>
<td>5.77</td>
<td>.84</td>
<td>.69</td>
<td>.51**</td>
<td>-.38**</td>
<td>.14**</td>
<td>.34**</td>
<td>.38**</td>
<td>-.34**</td>
<td>-.07*</td>
<td>.22**</td>
<td>-.47**</td>
<td>.44**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>12. Task Performance</td>
<td>5.99</td>
<td>.82</td>
<td>.83</td>
<td>.43**</td>
<td>-.31**</td>
<td>.16**</td>
<td>.34**</td>
<td>.31**</td>
<td>-.30**</td>
<td>-.07*</td>
<td>.09**</td>
<td>-.30**</td>
<td>.46**</td>
<td>.60**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* p < .05. **p < .01.
Table 2. MM-IRT-C Models (Final Model for each construct is italicized)

<table>
<thead>
<tr>
<th>Model</th>
<th>Model description</th>
<th>Log likelihood</th>
<th>BIC</th>
<th>CAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>Openness - 1 class</td>
<td>-8684.92</td>
<td>17658.29</td>
<td>17700.29</td>
</tr>
<tr>
<td>O2</td>
<td>Openness - 2 classes</td>
<td>-8655.10</td>
<td>17612.40</td>
<td>17656.40</td>
</tr>
<tr>
<td>O3</td>
<td>Openness - 3 classes</td>
<td>-8654.55</td>
<td>17625.03</td>
<td>17671.03</td>
</tr>
<tr>
<td>O4</td>
<td><em>Openness - 2 classes with Social Desirability only</em></td>
<td>-8649.44</td>
<td>17607.94</td>
<td>17652.94</td>
</tr>
<tr>
<td>O5</td>
<td>Openness - 2 classes with Depression only</td>
<td>-8652.40</td>
<td>17613.87</td>
<td>17658.87</td>
</tr>
<tr>
<td>A1</td>
<td>Agreeableness - 1 class</td>
<td>-8339.45</td>
<td>16974.00</td>
<td>17017.00</td>
</tr>
<tr>
<td>A2</td>
<td>Agreeableness - 2 classes</td>
<td>-8330.24</td>
<td>16969.31</td>
<td>17014.31</td>
</tr>
<tr>
<td>A3</td>
<td>Agreeableness - 3 classes</td>
<td>-8327.13</td>
<td>16976.81</td>
<td>17023.81</td>
</tr>
<tr>
<td>A4</td>
<td>Agreeableness - 2 classes with covariates</td>
<td>-8320.07</td>
<td>16962.69</td>
<td>17009.69</td>
</tr>
<tr>
<td>A5</td>
<td><em>Agreeableness - 2 classes with Social Desirability only</em></td>
<td>-8322.70</td>
<td>16954.22</td>
<td>16999.22</td>
</tr>
<tr>
<td>A6</td>
<td>Agreeableness - 2 classes with Depression only</td>
<td>-8397.67</td>
<td>17104.17</td>
<td>17149.17</td>
</tr>
<tr>
<td>N1</td>
<td>Neuroticism - 1 class</td>
<td>-8664.54</td>
<td>17624.71</td>
<td>17667.71</td>
</tr>
<tr>
<td>N2</td>
<td><em>Neuroticism - 2 classes</em></td>
<td>-8628.74</td>
<td>17566.87</td>
<td>17611.87</td>
</tr>
<tr>
<td>N3</td>
<td>Neuroticism - 3 classes</td>
<td>-8624.65</td>
<td>17572.44</td>
<td>17619.44</td>
</tr>
<tr>
<td>N4</td>
<td>Neuroticism - 2 classes with covariates</td>
<td>-8624.00</td>
<td>17571.14</td>
<td>17618.14</td>
</tr>
<tr>
<td>N5</td>
<td>Neuroticism - 2 classes with Social Desirability only</td>
<td>-8665.82</td>
<td>17641.02</td>
<td>17686.02</td>
</tr>
<tr>
<td>N6</td>
<td>Neuroticism - 2 classes with Depression only</td>
<td>-8697.31</td>
<td>17704.00</td>
<td>17749.00</td>
</tr>
<tr>
<td>C1</td>
<td>Conscientiousness - 1 class</td>
<td>-8688.12</td>
<td>17672.05</td>
<td>17715.05</td>
</tr>
<tr>
<td>C2</td>
<td>Conscientiousness - 2 classes</td>
<td>-8665.23</td>
<td>17640.04</td>
<td>17685.04</td>
</tr>
<tr>
<td>C3</td>
<td>Conscientiousness - 3 classes</td>
<td>-9140.66</td>
<td>18607.20</td>
<td>18654.20</td>
</tr>
<tr>
<td>C4</td>
<td><em>Conscientiousness - 2 classes with covariates</em></td>
<td>-8656.65</td>
<td>17636.63</td>
<td>17683.63</td>
</tr>
<tr>
<td>C5</td>
<td>Conscientiousness - 2 classes with Social Desirability only</td>
<td>-8676.17</td>
<td>17661.91</td>
<td>17706.91</td>
</tr>
<tr>
<td>C6</td>
<td>Conscientiousness - 2 classes with Depression only</td>
<td>-8689.77</td>
<td>17689.12</td>
<td>17734.12</td>
</tr>
<tr>
<td>E1</td>
<td>Extraversion - 1 class</td>
<td>-9248.53</td>
<td>18785.78</td>
<td>18827.78</td>
</tr>
<tr>
<td>E2</td>
<td>Extraversion - 2 classes</td>
<td>-9193.95</td>
<td>18690.36</td>
<td>18734.36</td>
</tr>
<tr>
<td>E3</td>
<td>Extraversion - 3 classes</td>
<td>-9181.89</td>
<td>18680.00</td>
<td>18726.00</td>
</tr>
<tr>
<td>E4</td>
<td>Extraversion - 4 classes</td>
<td>-9181.75</td>
<td>18693.47</td>
<td>18741.47</td>
</tr>
<tr>
<td>E5</td>
<td><em>Extraversion - 3 classes with Social Desirability only</em></td>
<td>-9151.88</td>
<td>18633.73</td>
<td>18681.73</td>
</tr>
<tr>
<td>E6</td>
<td>Extraversion - 3 classes with Depression only</td>
<td>-9181.08</td>
<td>18692.12</td>
<td>18740.12</td>
</tr>
</tbody>
</table>
Table 3. Correlations Between Personality Latent Classes and the Criterion (n = 947 - 1026)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Neuroticism class</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Conscientiousness class</td>
<td>.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Agreeableness class</td>
<td>.04</td>
<td>.09**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Extraversion class</td>
<td>-.03</td>
<td>.04</td>
<td>.04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Openness class</td>
<td>-.02</td>
<td>.03</td>
<td>-.02</td>
<td>-.04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. CWB</td>
<td>-.06</td>
<td>.11**</td>
<td>-.03</td>
<td>-.10**</td>
<td>.03</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. OCBI</td>
<td>.08**</td>
<td>.15**</td>
<td>.10**</td>
<td>.01</td>
<td>.05</td>
<td>-.25**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. OCBO</td>
<td>.07*</td>
<td>-.25**</td>
<td>-.03</td>
<td>.05</td>
<td>.02</td>
<td>-.47**</td>
<td>.44**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>9. Task Performance</td>
<td>.02</td>
<td>-.17**</td>
<td>.01</td>
<td>.05</td>
<td>.04</td>
<td>-.30**</td>
<td>.46**</td>
<td>.60**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* p < .05. **p < .01.
Figure 1. Mean Response Pattern for Openness classes, y-axis - probability of endorsement; x-axis - scale items
Figure 2. Mean Response Pattern for Agreeableness classes, y-axis - probability of endorsement; x-axis - scale items
Figure 3. Mean Response Pattern for Neuroticism classes, y-axis - probability of endorsement; x-axis - scale items
Figure 4. Mean Response Pattern for Conscientiousness classes, y-axis - probability of endorsement; x-axis - scale items
Figure 5. Mean Response Pattern for Extraversion classes, y-axis - probability of endorsement; x-axis - scale items