MEASUREMENT INVARIANCE OF BURNOUT INVENTORIES ACROSS SEX

Garett Coleman Foster

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

MASTER OF ARTS
May 2015

Committee:
Michael Zickar, Advisor
Steve Jex
William O'Brien
ABSTRACT

Michael Zickar, Advisor

This study investigated potential causes of extant group mean differences of self-reported levels of burnout across sex. Based on the origins of burnout as an interpersonal construct, combined with research showing differences in how men and women perceive and resolve interpersonal conflict and stress, it was hypothesized that statistical artifacts, namely violations of the measurement invariance assumption, were present. It was further hypothesized that these artifacts were causing differences in the prevalence of self-reported burnout and that removing them would reduce the size of such differences. Participants (n = 579) completed three separate measures of burnout: the Copenhagen Burnout Inventory, the Oldenburg Burnout Inventory, and the Maslach Burnout Inventory, as well as several measures of nomological correlates of burnout. The data were analyzed using confirmatory factor analysis and item response theory approaches to measurement invariance. Results indicated that no violations of measurement invariance were present. Implications for future research and practice are discussed.
I dedicate this thesis to my parents, Ron and Kim Foster,

for all the inspiration and encouragement they have given me.
ACKNOWLEDGMENTS

I would like to acknowledge and thank my committee, Drs. Zickar, Jex, and O’Brien, for their feedback on previous versions of this manuscript and their support of my research. I would also like to thank Rachel King for her guidance throughout the duration of this research. I couldn’t have done it without you all.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Burnout</td>
<td>1</td>
</tr>
<tr>
<td>Burnout and sex</td>
<td>5</td>
</tr>
<tr>
<td>Confirmatory Factor Analysis and Measurement Invariance</td>
<td>7</td>
</tr>
<tr>
<td>Item Response Theory and Differential Functioning</td>
<td>8</td>
</tr>
<tr>
<td>Ideal point response process</td>
<td>11</td>
</tr>
<tr>
<td>Differential Prediction</td>
<td>12</td>
</tr>
<tr>
<td>METHODS</td>
<td>14</td>
</tr>
<tr>
<td>Participants</td>
<td>14</td>
</tr>
<tr>
<td>Measures</td>
<td>14</td>
</tr>
<tr>
<td>Burnout measures</td>
<td>14</td>
</tr>
<tr>
<td>Maslach Burnout Inventory</td>
<td>14</td>
</tr>
<tr>
<td>Copenhagen Burnout Inventory</td>
<td>15</td>
</tr>
<tr>
<td>Oldenburg Burnout Inventory</td>
<td>15</td>
</tr>
<tr>
<td>Job satisfaction</td>
<td>16</td>
</tr>
<tr>
<td>Intention to leave</td>
<td>16</td>
</tr>
<tr>
<td>Physical symptoms</td>
<td>16</td>
</tr>
<tr>
<td>Interpersonal conflict</td>
<td>17</td>
</tr>
<tr>
<td>Workload</td>
<td>17</td>
</tr>
<tr>
<td>Interpersonal orientation</td>
<td>17</td>
</tr>
<tr>
<td>RESULTS</td>
<td>18</td>
</tr>
</tbody>
</table>
### Table of Contents

- Preliminary Analysis .................................................................................................. 18
- Confirmatory Factor Analyses ................................................................................... 19
  - Model fit......................................................................................................... 19
  - Measurement invariance ................................................................................ 20
- Item Response Theory ............................................................................................... 21
  - Item-level fit................................................................................................... 21
  - Differential item functioning ......................................................................... 21
  - Unfolding item response theory ..................................................................... 24
- Differential Prediction ............................................................................................... 25

#### DISCUSSION

- Limitations and Future Research ............................................................................... 29
- Conclusion ........................................................................................................... 30

#### REFERENCES ...................................................................................................................... 32

#### APPENDIX A ........................................................................................................ 41

#### APPENDIX B ............................................................................................................ 51
INTRODUCTION

Burnout has been studied by many researchers over the last 30 years, beginning with Maslach and Jackson in 1981. Starting with ethnographic interviews of real workers, Maslach and Jackson described burnout as a process of reacting to and coping with interpersonal stress. Though the definition has evolved over the years, burnout still has roots in interpersonal stress that can be seen by examining various measures of the construct. Spector and Jex (1998) defined interpersonal stressors at work as conflict between two people that can be either overt (e.g. face-to-face harassment) or covert (e.g. telling lies about someone behind their back). As will be discussed in detail below, several researchers have found sex differences in how interpersonal stress develops and is dealt with. Given these sex differences, in conjunction with the interpersonal origin of burnout, this could raise some concerns with how the construct and the measures of the construct operate across sex; this potential problem is the focus of this research.

Burnout

In their seminal 1981 article, Maslach and Jackson defined burnout as “a syndrome of emotional exhaustion and cynicism that occurs frequently among individuals who do ‘people-work’ of some kind” (p. 99); they also noted a third dimension of burnout, namely reduced personal efficacy, which is a self-evaluation of one’s inability to perform or be productive in their work. Twenty years later in an Annual Review of Psychology, Maslach, Schaufeli, and Lieter (2001) described burnout as “a prolonged response to chronic emotional and interpersonal stressors on the job” (p. 397), noting the same three dimensions. Others (e.g. Demerouti, Bakker, Vardakou, & Kantas, 2003; Gustavsson, Hallsten, & Rudman, 2010; Kristensen, Borritz, Villadsen, & Christensen, 2005), however, have disagreed with these three dimensions and have proposed alternative one- two- and three-factor conceptualizations of their own. Across these
and other debates about the operationalization of burnout, there have been several
commonalities. One of these commonalities, the focal point of this article, is the difference in
the prevalence of burnout between men and women. This study investigated this inequity from a
new angle through the assessment of measurement invariance. First, however, more explication
of burnout is warranted.

The concept of burnout can be found in social commentaries and anecdotal cases going
back many years, but it was not until the 1970s and 1980s that researchers began to take a
systematic and scientific look at it. As mentioned above, initial work was exploratory in nature
with the goal of describing the phenomenon as it occurred in the real world, with particular
emphasis on occupations with strong interpersonal aspects such as service occupations (Maslach,
Schaufeli, & Leiter, 2001). It is from this body of research that the three-dimensional model of
burnout emerged, which consisted of emotional exhaustion, cynicism, and reduced professional
efficacy. Maslach and Jackson (1981) used these dimensions as the basis for the Maslach
Burnout Inventory (MBI), which has gone through several iterations since its first version to
create versions designed specifically for human services occupations and for teachers and
educators. In the 1990s, as interest and research on burnout began to expand, additional forms
were created for occupations outside of those with strong interpersonal components (e.g.
computer technicians), notably the General Survey (MBI – GS, Schaufeli, Leiter, Maslach, &
Jackson, 1996). Since its introduction, the MBI in its various forms has dominated the field of
research on burnout, with Schaufeli and Enzmann (1998) reporting that upwards of 90% of
dissertation research done on burnout between 1976 and 1996 used the MBI in one form or
another.
Despite the dominance of a single measure of burnout, there is still a great deal of debate regarding the construct. In 2005, the journal *Work and Stress* published a special edition featuring articles regarding the conceptualization and measurement of burnout to address this debate. This edition featured two new burnout measures that have different conceptualizations of the construct than the MBI.

The first of these two measures was the Copenhagen Burnout Inventory (CBI) proposed by Kristensen, Borritz, Villadsen, and Christensen (2005). In laying out their rationale for a new measure of burnout, the authors described several issues they found with the MBI. Notable among these arguments was the notion that the three dimensions of Maslach’s model of burnout represent three fundamentally different types of constructs. Specifically, they felt that emotional exhaustion represents an individual state resulting from environmental effects; that cynicism represents a coping strategy in response to emotional exhaustion, which Maslach has been quoted as supporting (Maslach, 1979; as cited in Kristensen, Borritz, Villadsen, and Christensen, 2005); and that reduced personal efficacy is an outcome of the previous two dimensions. As such, they created their instrument to measure only emotional exhaustion, which they and others (e.g. Bekker, Croon, & Bressers, 2005) have felt is the core concept of burnout. The CBI is divided into three separate subscales which measure personal burnout, work-related burnout, and client-related burnout. The first subscale, personal burnout, is designed to be a measure of generic fatigue or exhaustion, and the authors note that the term ‘burnout’ is only used to designate the scale as part of the CBI. The second subscale, work-related burnout, is designed to measure burnout among workers of any occupation regardless of the nature of their work. The third subscale, client-related burnout, pertains specifically to those workers who deal with “people work” (p. 197).
The second recently developed burnout measure, from Halbesleben and Demerouti (2005), is an English language adaptation of the Oldenburg Burnout Inventory (OLBI, Demerouti, Bakker, Vardakou & Kantas, 2003). In contrast to the three dimensions of the MBI and the one dimension of the CBI, the OLBI is comprised of two dimensions, exhaustion and disengagement (an equivalent counterpart to cynicism on the MBI). The authors note that a large advantage that the OLBI has over the MBI is that the OLBI is designed specifically to measure both physical and cognitive exhaustion in employees, whereas the MBI is solely cognitive. Both this measure and the CBI offer ways of measuring alternative conceptualizations of burnout, which has been of great interest to researchers for some time (Shirom, 2005). In spite of the theoretical advances offered by these new conceptualizations, sex differences in burnout persist in the literature.

Burnout also has been linked to several antecedents and consequences that have important implications for employers and employees alike. Maslach, Schaufeli, and Lieter (2001) stated that burnout’s nomological network includes variables such as workload and interpersonal conflict as predictors as well as reduced job satisfaction, greater withdrawal intentions, and the presence of physical symptoms such as headaches and fatigue as outcomes. Numerous studies (e.g. Brackett, Palomera, Mojsa-Kaja, Reyes, & Salovey, 2010; Faller, Gates, Georges, & Connelly, 2011; Jayaratne & Chess, 1984; Weisberg, 1994) have corroborated the reports of Maslach and colleagues using a variety of different samples and shown that the detriment that burnout can have on an organization can be substantial. As will be discussed later, however, attempting to relate these variables to burnout using measures containing significant measurement artifacts (e.g. differential item functioning) could result in undesirable psychometric properties.
Burnout and sex.

Sex differences in the prevalence of burnout have been studied to some degree of emphasis or another by many of researchers (e.g. Ahola et al., 2006; Maslach, Schaufeli, & Leiter, 2001). Most often, women have been found to experience slightly higher levels of burnout than men (Gleichgerrcht & Decety, 2013; Cheng, Chen, Chen, Burr, & Hasselhorn, 2013), but this is not always the case. For instance, Schaufeli and Enzmann (1998) found that whereas women had higher levels of emotional exhaustion, men had higher levels of depersonalization. Bekker, Croon, and Bressers (2005) also found that men had higher emotional exhaustion. A meta-analysis by Purvanova and Muros (2010) found that, overall, women experienced more overall burnout than men and that this relation was not moderated by occupation \( (d = .18) \). These differences in burnout between sex have been found to hold across occupation (Guthrie & Jones, 2012), nationality (Gbadamosi & Ross, 2012), and level of education (Ahola et al., 2006). None of these studies, however, were able to determine the root cause of these differences.

It has been suggested that sex differences in burnout occur because of higher empathetic concern among women (Gleichgerrcht & Decety, 2013) or because women are more emotionally expressive (Purvanova & Muros, 2010). However, the cause of these group differences in burnout may lie at a more fundamental level. Several studies examined the underlying processes involved in stress across sex. Barbosa-Leiker and colleagues (2013) found that men and women have different experiences of both stress and counter-stress, which the authors operationalized as negatively-worded stress items. Using a confirmatory factor analysis approach, these authors found lower latent means for stress among men and higher latent means of counter-stress. Additionally, Bouchard and Shih (2013) found out that women perceive and generate interpersonal stress differently than do men using a time-lagged hierarchical modeling approach.
These authors suggested that the sex differences they found in interpersonal stress may be due to women having more interpersonal orientation, as shown by others (Cross & Madson, 1997; McBride, Bacchiochi, & Bagby, 2005). They also noted that neediness and co-rumination only partially mediated the gender-stress relation, leaving sex as a significant factor. These findings could be important for understanding burnout differences given burnout’s origins as a reaction to interpersonal stress.

Several studies have also examined sex and stress differences specifically in the workplace. Nag, Poddar, and Nag (2005) found that men and women perceived aspects of work as stressors in significantly different ways (as cited in Herreo, Saldaña, Rodrigues, & Ritzel, 2012). Watson, Goh, and Sawang (2011) conducted separate path analyses for men and women and found that these groups have different processes for reacting to and dealing with stress in the workplace, which is useful to distinguish since emotional exhaustion and cynicism (or their equivalents on other scales) are a reaction and a coping mechanism, respectively. Because of these differences, it is possible that items written to measure burnout may take on different meanings for women and men, specifically those items related to interpersonal stress.

A better understanding of the sex differences in burnout is important for employees of both sexes. Maslach, Schaufeli, and Leiter (2001) report that some believe that “burnout is more of a female experience” (p. 410). As Purvanova and Muros (2010) noted, perceptions of women experiencing more burnout are harmful to both sexes for two reasons: 1) this may lead to prejudices that women cannot handle the stress related to higher level jobs, which could lead to qualified women not receiving promotions or assignments, and 2) this belief may lead to real burnout in men being discounted as non-existent, which would lead to these men not receiving important help and interventions. I believe that examining burnout using the methods described
below, known collectively as measurement invariance research, will help address these practical issues related to burnout. Additionally, this research can help shed new light on the theoretical underpinnings of burnout and how this phenomenon functions while opening up new avenues of research in the area.

**Confirmatory Factor Analysis and Measurement Invariance**

Confirmatory factor analysis (CFA), a method of structural equation modeling, can be used to assess group differences by examining the assumption of measurement invariance. Measurement invariance is the idea that the psychometric properties of a scale or its constituent items are equivalent (invariant) when assessed within different groups individually. This is important to researchers because violations of this assumption across groups can make comparison of the scale scores of those groups difficult or impossible and interpretation questionable.

Vandenberg and Lance (2000) offer a comprehensive review of current practices in CFA measurement invariance research and offer guidelines for conducting such research. These authors showed that procedures are not wholly universal and that different authors use slightly different techniques to best suit their research questions. To address issues of group mean differences, which is the focus of this study, researchers should establish total measurement invariance, which subsumes tests of configural invariance (equivalent factor structures), metric invariance (equivalent factor loadings), scalar invariance (equivalent factor intercepts), and a test of equivalent error variances. These tests, which are essentially a process of successively more restrictive equality, are considered sequential and should thus only be conducted if the previous test demonstrated invariance. Based on these principles, I sought to answer the following research question:
Research Question 1: Will confirmatory factor analyses demonstrate sufficient measurement invariance to allow for cross-group comparisons?

**Item Response Theory and Differential Functioning**

Item response theory (IRT) is “a system of models that defines one way of establishing the correspondence between latent variables and their manifestations” (De Ayala, 2009, p. 4). “Latent variables” refers to underlying constructs such as cognitive ability, personality, or levels of stress. The “manifestations” are the items on a test or scale designed to measure the underlying construct. The fundamental concept underlying IRT is that the likelihood of an individual endorsing an item is a function of where both the individual and the item are located on the latent trait continuum being measured (Drasgow & Hulin, 1990). An individual’s underlying trait is given by the Greek letter theta (θ) and an item’s location on the same continuum is given by the letter b. A key difference between IRT and classical test theory (CTT) is that CTT looks only at the individual’s score on the test as a whole, whereas IRT allows analysis of how individual items function to estimate a person’s trait score. Additionally, whereas CTT assumes a linear relation between the latent construct and a person’s response, IRT assumes a more flexible relation. This allows researchers to examine how different items predict or model an individual’s measured traits. The relation can be represented graphically by an item response function (IRF), which shows the probability of individuals across all levels of θ endorsing a given item.

The IRT model used here to analyze the data is Samejima’s (1969) graded response model (GRM). The GRM, which is used for polytomous data such as data from a Likert-type response scale, has two item parameters in addition to the individual latent trait parameter θ: the discrimination parameter given by the letter a, which represents how well the item differentiates
between individuals located near the item’s location on the latent continuum; and the threshold or boundary parameter given by the letter \( b \), which represents where an individual would need to be located on the latent continuum in order to endorse a specific response to the item. Note that the \( b \)-parameter corresponds to the item location parameter discussed above. For each item with \( k \) response categories, there are \( k - 1 \) boundary parameters. For example, items of the Oldenburg Burnout Inventory are rated on a 1 to 4 scale, so there are 4 response categories and therefore 3 boundary parameters for each item.

Differential item functioning (DIF), which is a complimentary form of measurement invariance assessment to the CFA approach (Raju, Lafitte, & Byrne, 2002), is a method of analysis that uses statistical models based in IRT to detect differences in the performance of items across groups (De Ayala, 2009). These differences are reflected in different values for the item parameters \( a \) or \( b \) when they are calculated for each group separately. Parsing out these item properties from respondent effects allows researchers to determine if observed group mean differences are the result of true differences in the construct or an artifact of biased items. Differing \( b \) parameters result in uniform DIF, which reflects items that are easier for one group to endorse than for another equally across all levels of \( \theta \). Differing \( a \) parameters result in non-uniform DIF and reflects items that are more or less difficult for one group to endorse than the other at varying levels of \( \theta \). This study will focus on identifying both types of DIF across gender.

Differential test functioning (DTF) is an aggregation of the effects of DIF across the entirety of a scale or test. This is a useful analysis because different items may exhibit DIF in different directions (i.e. one item favors one group and a second item favors the other group). In some cases (e.g. Drasgow, 1987), many items can exhibit DIF but the overall test effect will
result in no difference because the number of items favoring the focal group are balanced out by an equal number of items favoring the reference group. In addition to studying DIF, this study addresses the practical implications of item bias by investigating DTF.

IRT analysis is relatively uncommon in the burnout literature. Only two studies utilizing this methodology could be located. Gonzalez-Roma, Schaufeli, Bakker, and Lloret (2006) examined the structure of how work engagement, burnout’s proposed conceptually opposite construct, relate to one another using Mokken scaling, a non-parametric IRT model used for small sample sizes to test for unidimensionality of a construct, and did not examine differences between any groups. Gustavsson, Hallsten, and Rudman (2010) used the Rasch IRT model, a very simplified and restrictive model (Andrich, 1988a,b; Wright & Mok, 2000) to investigate burnout as a process occurring across time. This study also examined how items of the Oldenburg Burnout Inventory (OLBI, Demerouti, Bakker, Vardakou, & Kantas, 2003) operated across individuals of different groups. They found some evidence of both forms of DIF across groups of younger and older workers for several items; their study is the only examination of any type of DIF in the burnout literature that could be found. Neither study, however, examined sex differences.

Because of the mathematical relation between an item’s location on the latent continuum and that parameter’s effect on an individual’s observed score, I propose the following hypotheses:

Hypothesis 1a: Items related to interpersonal stress will exhibit differential item functioning such that the discrimination parameters will vary across sex.

Hypothesis 1b: Items related to interpersonal stress will exhibit differential item functioning such that the threshold parameters will be lower for women than for men.
Hypothesis 1c: Scales with differentially functioning items will exhibit differential test functioning such that the test characteristic curve for women will be lower than the test characteristic curve for men.

Ideal point response process.

Like many scales in use today, the burnout inventories described above are scored using Likert-type response scales. Responses to this type of scale often follow a dominance response process (Coombs, 1960). According to Roberts, Laughlin, and Wedell (1999), an individual following a dominance response process will endorse an item “to the extent that the individual is located above the item on the underlying continuum” (p. 215) which is analyzed with a cumulative probability model. The issue with this is that several researchers (e.g. Andrich, 1996; Roberts, 1995) have found that individuals in an empirical setting use an ideal point response process for attitude statements (such as those comprising the burnout inventories). In an ideal point response process, an individual’s likelihood of endorsing an item is given as a function of the distance between the extremeness of the item and the extremeness of the person’s attitude on the latent continuum being measured (Coombs, 1960). In other words, if an individual is extremely high on the latent continuum, their probability of endorsing an item will actually go down if the item is not extreme enough, which is in contrast to the dominance response process assumed by Likert-type scales that would expect an individual’s likelihood of endorsement to monotonically increase.

Ideal point response process assumptions have been used with success in industrial/organization psychology (e.g. Stark, Chernyshenko, & Drasgow, 2006). However, no such studies could be found specifically within the burnout literature. As such, it may be very beneficial to check for improved model fit using the Generalized Graded Unfolding Model
(Roberts, Donoghue, & Laughlin, 1998), an IRT model designed specifically for ideal point response systems. Additionally, Wang, Tay, and Drasgow (2013) developed procedures to investigate differential item and test functioning using the GGUM, which can be used if DIF is detected under the GRM. These possibilities give rise to the following research question:

Research Question 2a: Will analyzing burnout data using an ideal point response process approach result in improved model fit?

Research Question 2b: Will analyzing burnout data using an ideal point response process approach eliminate observed differential functioning of the inventories?

**Differential Prediction**

Stark, Chernyshenko, and Drasgow (2004) defined differential prediction as “the notion that tests do not predict equally well across groups.” Specifically, it refers to regression lines relating a construct to other constructs in its nomological network that have different slopes or intercepts when calculated separately for two different populations. Such an analysis can be easily accomplished using multiple hierarchical regression as described in Frederickson and Melville (1954). Despite the ease of such an analysis, differential predictability can have important legal and ethical consequences for organizations and research implications for scientists (Stark, Chernyshenko, & Drasgow, 2004). As such, it is an important tool to help inform the actual end-user effect of measurement invariance analysis. Given this basis and the possibility of DIF, I propose the following hypotheses:

Hypothesis 2: Scales exhibiting differential item or test functioning will also exhibit differential predictability in the relations with burnout’s antecedents and consequences.

Hypothesis 3: Differential predictability will be reduced by removing items that exhibit differential item functioning.
Given the scarcity of measurement invariance research in the field of burnout, this study helps to fill a gap in the literature. As alluded to above, these analyses offer a much deeper examination and understanding of how people view items and the constructs that those items measure. This is particularly beneficial to burnout due to the current debate surrounding the conceptualization and measurement of the construct. In addition to addressing this gap, one other goal of this research is to open up the field of burnout research to a more thorough and rigorous application of IRT methods. Previous studies, although constructive, used limited models and did not consistently conduct a meticulous assessment of model fit, which is central to the use of IRT.
METHODS

Participants

Participants for this study were gathered using Amazon’s Mechanical Turk service (https://www.mturk.com/mturk/). Mturk has been shown to produce reliable data comparable to traditional data collection methods, with some evidence that samples obtained this way may represent more diverse populations (Buhrmester, Kwang, & Gosling, 2011). Participants were paid one dollar to fill out the 10 minute online survey. The survey included several validation check items (e.g. Please respond “Agree” to this item) to check for insufficient effort in filling out the questionnaire. Forty-four responses failed these checks and were thrown out, with an additional 34 removed for taking less than five minutes to complete the 134 item survey. The final sample of 579 participants was 60% male, with 50% of respondents falling into the 24-34 years of age category and 81% having at least 12 months of experience or more at their current job.

Measures

The scales used in this study can be found in Appendix B. The internal consistencies (α) for each scale can be found in parentheses on the diagonal of table 1 (all α ≥ .69).

Burnout measures.

Maslach Burnout Inventory.

The Maslach Burnout Inventory General Survey (MBI-GS, Schaufeli, Leiter, Maslach, & Jackson, 1996) is an adapted form of the original measure designed to be used with any population of workers regardless of occupation. The total measure is comprised of 16 items with six items pertaining to personal efficacy and five items for both the exhaustion and cynicism
scales. Item ratings are on a frequency scale and are measured from 0 (Never) to 6 (Everyday).

None of the items are reverse scored, and higher scores on all three subscales are indicative of higher burnout. It is important to note that this is not usually the case for the Personal Efficacy subscale; higher scores usually indicate less burnout, but scores were reversed in this study to simplify interpretation of scores. An overall burnout score is not computed across subscales, as is the case with the other burnout inventories. As the MBI-GS is a proprietary scale, it does not appear in Appendix B. The Cronbach’s Alpha for the Emotional Exhaustion, Cynicism, and Personal Efficacy subscales of the MBI were .92, .86, and .79, respectively, indicating good internal consistency of the scales.

*Copenhagen Burnout Inventory.*

The Copenhagen Burnout Inventory (CBI, Kristensen, Borritz, Villadsen, and Christensen, 2005) is a 19-item measure subsuming three separate scales: personal burnout, work-related burnout, and client-related burnout. Items are measured using five-category responses. The specific categories are *Never/Almost never, Seldom, Sometimes, Often,* and *Always* for 12 of the items and *To a very low degree, To a low degree, Somewhat, To a high degree,* and *To a very high degree* for the other 7 items. Only one item, which appears on the work-related burnout scale, uses reverse scoring. The CBI is made free for academic use. The internal consistency for the Personal, Work, and Client subscales of the CBI were .83, .88, and .91, respectively.

*Oldenburg Burnout Inventory.*

The Oldenburg Burnout Inventory (OLBI, Demerouti, Bakker, Vardakou, & Kantas, 2003) is a 16-item measure with two subscales measuring exhaustion and disengagement, each being measured by eight items. Items are measured with four response categories (1 = *Strongly
Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree). Of the eight items on each subscale, four items use reverse scoring. The OLBI is also free for academic use. The Disengagement and Exhaustion subscales had internal consistency scores of .82 and .84, respectively.

**Job satisfaction.**

Job satisfaction is measured using the Job Descriptive Index (JDI). Originally developed by Smith, Kendall, and Hulin (1969), the JDI has undergone several iterations and revisions (Kihm, Smith, & Irwin, 1997; Smith, et al., 1987) to examine and improve its psychometric qualities. The current version (Lake, Gopalkrishnan, Sliter, & Withrow, 2010), consists of six subscales, including the Job In General (JIG) scale. Participants respond to each item with Yes, No, or “?” as it relates to their job. The response options are then scored as 3, 0, or 1, respectively. Because the questionnaire for this survey was quite long and the focus of the study is on interpersonal interactions, only the Coworker subscale of the JDI and the JIG are measured here. The Coworker subscale had a Chronbach’s Alpha of .93, and the JIG had an Cronbach’s Alpha of .92.

**Intention to leave.**

Withdrawal intentions were measured using the Intention to Quit Scale (Crossley, Grauer, Lin & Stanton, 2002; as cited in Crossley, Bennett, Jex, & Burnfield, 2007). This 5-item scale is measured using a Likert-type response from 1 (Strongly Disagree) to 7 (Strongly Agree). One item of this scale is reverse coded. This scale had a very high internal consistency (α = .97), indicating potential redundancy among items.

**Physical symptoms.**

Participant’s physical symptoms were measured using self-report via the 13-item version of the Physical Symptoms Inventory (Spector & Jex, 1997). This scale assesses how often
during the past month the individual experiences mild to moderate physical issues (e.g. nausea, headache, fatigue) using a five-point scale (1 = Not at all, 2 = Once or Twice, 3 = Once or Twice per Week, 4 = Most Days, 5 = Every day). The PSI had an internal consistency score of .89.

**Interpersonal conflict.**

To assess participants’ levels of interpersonal conflict at their job, the Interpersonal Conflict at Work Scale (ICAWS; Spector & Jex, 1997) was used. This four-item measure is rated on a 1 (Never) to 5 (Very Often) response scale and assesses the frequency with which an individual fights with or is victimized at work. The internal consistency for the ICAWS was .85.

**Workload.**

Workload was measured using the five-item Quantitative Workload Inventory (Spector & Jex, 1997). The responses to each item (1 = Less than once per month or never, 2 = Once or twice per month, 3 = Once or twice per week, 4 = Once or twice per day, 5 = Several times per day) reflect how often an individual’s workload is heavy enough to be considered stressful. The QWI had a Cronbach’s Alpha of .84.

**Interpersonal orientation.**

Finally, as a manipulation check, participant’s interpersonal orientation was measured using the Interpersonal Orientations Scale (IOS; Swap & Rubin, 1983). This scale consists of 29 items measuring a person’s orientation toward interpersonal interactions. Four items were removed for content relevance (e.g. I find myself wondering what telephone operators are really like). The final scale was scored on agreement from 1 (Strongly Disagree) to 5 (Strongly Agree) with seven reverse-coded items. The IOS had a somewhat low but generally acceptable level of internal consistency (α = .69) for a scale of this length.
RESULTS

Preliminary Analysis

Means, standard deviations, and correlations of the study variables are presented in table 1. Note that all scores on the burnout inventories were coded so that higher scores reflect more burnout. Most participants showed mild to moderate levels of burnout. Of the 8 subscales that comprise the Copenhagen, Oldenburg, and Maslach Burnout Inventories, only 4 showed significant differences in burnout between men and women. These were the Personal ($d = -.32$) and Work ($d = -.30$) Burnout subscales of the CBI and the Emotional Exhaustion subscales of the OLBI ($d = -.30$) and MBI ($d = -.19$). It does warrant note that the Personal Burnout subscale of the CBI is a measure of general stress meant as a comparative baseline when needed and not as a true measure of burnout (Kristensen, Borritz, Villadsen, and Christensen, 2005). As mentioned above, all subscales of the CBI were designed to reflect types of emotional exhaustion with different targets. Thus, a clear pattern emerged from these data with women reporting higher levels of burnout than men at moderate effect sizes, corroborating past findings. Only 3 other variables measured showed differences across sex; these were physical symptoms ($d = -.28$), workload ($d = -.22$), and interpersonal orientation ($d = -.36$), with interpersonal orientation also supporting the groundwork for my subsequent hypotheses. Group means, $t$-statistics, and effect sizes for these 7 variables are presented in table 2.

Before examining measurement invariance, 14 doctoral student and faculty subject matter experts (SMEs) in Industrial/Organizational psychology were recruited to examine the items in each of the burnout inventories to determine if the items were interpersonal in nature. Each SME read each item and marked ‘Yes’ or ‘No’ to indicate if he/she believed the item to be interpersonal in nature. Items that 70 percent or more of the SMEs marked ‘Yes’ were
considered to be interpersonal. Only 7 of the 51 items met this criteria. Of these items, 6 coincidentally comprised the entirety of the Client-Related Burnout subscale of the CBI and all directly referenced clients (items were not labeled as belonging to particular scales, so it is unlikely that the name of the scale influenced any SME analyses). The seventh item, which referenced leisure time with friends and family, fell on the Work-Related Burnout subscale of the CBI. The relative lack of interpersonal items in the inventories throws into question the proposition that it is differences in interpersonal orientation that leads to differences in burnout. This notion will be further tested below.

**Confirmatory Factor Analyses**

All confirmatory factor analyses (CFA) were run using Mplus 7.2 (Muthén & Muthén, 2012).

**Model fit.**

Before testing for measurement invariance, model fit was established for each of the 8 subscales for the entire sample. Five of the subscales were found to fit the data well; the 3 scales that did not were the Disengagement scale of the OLBI ($\chi^2 (20) = 233.57, p < .001$, RMSEA (.123, .155), CFI = 0.88, SRMR = .064) and the MBI subscales Personal Efficacy ($\chi^2 (9) = 191.98, p < .001$, RMSEA (.167, .214), CFI = 0.83, SRMR = .076) and Cynicism ($\chi^2 (5) = 234.33, p < .001$, RMSEA (.254, .316), CFI = 0.87, SRMR = .057). Modification indices were examined for psychometric practicality and theoretical justifiability. Freeing the error correlation between items 4 and 5 on the MBI Cynicism scale drastically improved model fit ($\chi^2 (4) = 11.05, p = .026$, RMSEA (.017, .096), CFI = 0.996, SRMR = .018). Examination of the content of these items revealed that they were more similar to one another than to the other 3 items on the scale (i.e., they both referred to intense positive emotions). As such, the
modification was kept and all CFA measurement invariance analyses were run using the modification. No practical or theoretical modifications emerged for the other 2 subscales; measurement invariance results for these subscales should be interpreted with caution (Vandenberg & Lance, 2000).

**Measurement invariance.**

All CFA measurement invariance analyses were run according to the procedures laid out in Vandenberg and Lance (2000). Each successive test of configural, metric, and scalar invariance and invariant error variances represents the relaxing of an equality constraint across groups. These are a relaxation of factor loadings, item slopes, item intercepts, and item-level error terms, respectively. The tests are performed in order and tested for significance using a change in chi-squared test, with statistically significant changes representing a violation of measurement invariance. Because chi-squared tests are prone to influence from large samples, significance should be interpreted in the context of other changes of relative fit (Schmitt & Kuljanin, 2008).

Seven of 8 subscales were found to be invariant at each level of testing \((p > .05)\), showing that these scales exhibit measurement invariance across sex. The eighth scale, the Disengagement subscale of the OLBI, showed significance when relaxing the cross-group equality constraints on the item slopes \(\Delta \chi^2 (8) = 18.16, p = .02\). This would suggest a violation of metric invariance. However, the significance of this test was not corroborated by any other changes in fit statistics. Specifically, the CFI change was less than .01 and as such failed to meet the criterion laid out in Cheung and Rensvold (2002). In light of this and the fact that no items on this subscale were found to be invariant by the item response theory methods discussed below, testing on this subscale continued, with no further violations emerging. Thus, from a
CFA perspective, all 3 burnout inventories were found to be invariant, answering research question 1.

**Item Response Theory**

All traditional item response theory (IRT) analyses were conducted using Samejima’s (1969) Graded Response Model (GRM), which is capable of handling polytomous responses. Primary analyses were run in IRTPRO 2.1 (Cai, Thissen, & du Toit, 2011).

**Item-level fit.**

Table 3 presents fit statistics for all items of the 3 inventories. Fit was assessed using chi-squared statistics, with significant values indicating poor fit. Due to the large samples sizes required for IRT analyses, items with significant chi-squared values were assessed with a chi-squared to degrees of freedom ratio. A ratio of 3 or less ($\chi^2/df < 3.0$) is indicative of acceptable with, whereas ratios of less than 2 are considered excellent fit (Drasgow & Hulin, 1990). Using these cut-offs, all items were found to be well-fit by the GRM.

**Differential item functioning.**

To test for differential functioning of an item or set of items on a scale, it is necessary to select “anchor” items that are not suspected of differential item functioning (DIF) for use in linking responses across groups. Anchor items are chosen *a priori* based on reasonable belief that the item will not exhibit DIF. Items that were suspected of DIF were not used as anchors and therefore had their parameters equated across groups so that all output was on the same scale, allowing for comparison.

The original intent in this project was to use items not determined to be interpersonal as anchor items. Unfortunately, results of the SME analysis of the items resulted in 2 issues: 1) of the 7 items determined to be interpersonal in nature, 6 of them fell on the same subscale and
accounted for all items on that subscale, thus preventing any items from being used as anchors, and 2) 6 of the 8 subscales had no items labeled as interpersonal, so the original basis for a priori suspicion of DIF had no founding. In these instances, it is still possible to run DIF analyses using an iterative process wherein the software alternately anchors all items and tests all items. In this process, individual items are isolated and considered to be suspected of DIF, while all other items on that scale are used as anchors. DIF statistics are calculated for the item, and then the software anchors that item and tests a different item for DIF. This process continues until all items have been tested. This process has been used successfully to evaluate items that may or may not be suspected of DIF, such as during scale development (e.g. Chuah, Drasgow, & Roberts, 2006).

Because only 1 item would be testable for DIF based on the SME analysis and a priori anchors method, each scale was instead run using the anchor-all test-all method, thus allowing for the testing of the item in question as well as for any other item that may exhibit DIF for other reasons. Before proceeding, item fit was checked for both subgroups, and the data for all items were found to fit the model well in each. IRTPRO yields total and specific chi-squared statistics for each item, with significant values being indicative of DIF. Items are first examined at the total level. If significance is found here, then the specific (differences of the discrimination and/or location parameters) significance is examined to determine whether the DIF is uniform, non-uniform, or both in nature (Tay, Meade, & Cao, 2015).

Of the 51 items that comprise the 8 subscales of the 3 burnout inventories, only 2 were flagged by IRTPRO as exhibiting significant DIF. Both items fell on the Cynicism subscale of the MBI, with one item exhibiting uniform DIF (differences in the location parameters of the
items across groups) and one exhibiting non-uniform DIF (difference of the discrimination parameter across groups). These results are displayed in table 4.

Because the statistics used to determine significance of DIF can be unduly influenced by sample size, VisualDF (Meade, 2010) was used to examine the effect sizes of the differences. Before VisualDF could be used, subgroup parameter estimates had to be manually linked due to IRTPRO’s inability to export linked parameter estimates. This was accomplished through IRTEQ (Han, 2008) utilizing Mean/Sigma equating. Meade (2010) laid out a taxonomy of effect sizes to be used for IRT measurement invariance, all of which are outputted by VisualDF. Different indices were created to be useful in different situations. The two most pertinent to the current question of mean differences across groups are the Normed Signed Item Difference in the Sample (Normed SIDS) and the Normed Signed Test Difference in the Sample (Normed STDS). These indices are useful for assessing the practical impact of differential functioning at the item and test level, respectively. They allow for compensatory cancelation of effects, such that if different groups are favored at different times, the overall effects will cancel out, which is the case in observed score summation. Each of these effect sizes is measured in terms of standard deviations. Focus for assessing DIF (as opposed to DTF) should be given to the Normed SIDS, which is presented in table 4 alongside significance tests for the 2 items. The magnitude of these effect sizes is small enough to be considered negligible, thus showing that the impact of the DIF in these items is null. These effects are displayed graphically in figures 1 and 2. These graphs show the compensatory effect of the DIF, with men being favored for some response categories and ranges of $\theta$, whereas women are favored elsewhere. The Normed STDS was also nugatory (-0.013), showing that what little DIF was occurring had almost no effect at the subscale-score level. Given that the only 2 items found to exhibit DIF had nugatory effects, and that the effect
was all but washed out at the scale level, all subscales can be considered invariant from the IRT perspective. These findings do not support hypothesis 1.

Unfolding item response theory.

Item-level data were also examined from an ideal point response methodology to see if model fit could be improved with the use of unfolding. The Generalized Graded Unfolding Model (GGUM; Roberts, Donoghue, & Laughlin, 2000) was used because it is capable of handling polytomous responses and can be estimated using GGUM2004 (Roberts & Cui, 2004). The software automatically removes any respondents with zero variance in scores, resulting in slightly different sample sizes for each subscale analysis ($N = 559-579$). Additionally, 5 items resulted in a singularity (preventing computation of item information and fit) when analyzed and were unable to be examined from the unfolding perspective. The fit statistics for the remaining 46 items are found in table 4. The same rules of thumb for fit under the GRM also apply to the GGUM. In general, item fit was not improved through the use of the GGUM, with 4 items having marginal fit and 3 items with poor fit. This is not surprising when assessing the item response functions (IRFs) for the GGUM, which displayed little to no unfolding in any items along the typically observed range of theta scores (see figure 3 for an example). Figures 4 and 5 show a comparison of IRFs for the first item of the CBI Work Burnout subscale under the GRM and GGUM, respectively. The fact that these figures are very similar is indicative of the fact that fit was not improved with the more complex model. This conclusion answers research question 2a. Additionally, because no substantial DIF was detected under the dominance approach and no fit improvement was found, research question 2b is rendered unworkable.
Differential Prediction

Hypotheses 2 and 3 were predicated on the presence of DIF in the burnout inventories. As originally presented, these hypotheses cannot be supported. However, to ensure that the pattern of acceptable psychometric properties holds under a classical test theory approach, differential prediction was assessed for all antecedents and consequences of burnout across each of the subscales.

Before proceeding, each burnout subscale was regressed onto interpersonal orientation to assess whether variance in burnout could be explained by a person’s interpersonal orientation. None of the regression coefficients were significant after controlling for Type I error using a Bonferroni correction, and the variance explained was less than 1% for all scales. Thus, interpersonal orientation was found to not predict burnout and was not controlled for in the differential prediction analyses.

Differential prediction was assessed using hierarchical regression as laid out by Fredericksen and Mehlville (1954). Predictor variables and gender were entered as a block then an interaction term between the two was entered. A significant interaction term would be indicative of differential predictability. Of the 48 regressions run (8 burnout subscales with 6 variables from burnout’s nomological network), none resulted in a significant interaction. Thus no evidence of differential predictability was identified, further supporting the non-biased measurement properties of the burnout inventories.
DISCUSSION

Differences across sex have been prevalent in workplace burnout research for many years. These differences most often suggest that women are more burned out than men (Purvanova & Muros, 2010), and they have real, detrimental impacts on both sexes (Maslach, Schaufeli, & Lieter, 2001). Though the underlying root of these differences is not agreed upon, one possible cause is the difference in how men and women handle interpersonal situations (e.g. Bouchard & Shih, 2013; Cross & Madson, 1997; McBride, Bacchiochi, & Bagby, 2005). This study set out to address whether these differences led to statistical artifacts, specifically violations of measurement invariance, which may be causing the observed group mean differences.

The results of this study corroborated past research that found higher levels of emotional exhaustion, widely considered the core of burnout (Gustavsson, Hallsten, & Rudman, 2010), in women than in men across all three burnout inventories. Additionally, women reported higher mean levels of interpersonal orientation than did men. These effects were small to medium in size, providing a basis for supporting the notion that interpersonal orientation may lead to measurement invariance that in turn causes differences in burnout. However, interpersonal orientation failed to explain any variance in levels of burnout for either sex. Additionally, investigations of measurement invariance from both confirmatory factor analysis (CFA) and item response theory (IRT) perspectives failed to detect any psychometric properties of the scales that substantially differed across sex.

These results suggest that the empirical group mean differences in burnout are true group differences and not caused by statistical artifacts. Women do experience greater levels of burnout than do men. This is not, however, definitive evidence, despite of the methods used to
assess these hypotheses. There is some misbelief that all IRT analyses are unaffected by the sample from which the data are gathered. However, this is untrue. As such, this single study, while successful in moving this research area forward, is by itself insufficient to make final determinations on the sources of observed sex differences in burnout. Other researchers are encouraged to utilize similar methods to address this issue. Unfortunately, these results do little to shed additional light on the underlying causes of the group mean differences. As stated above, interpersonal orientation did not explain any variance in burnout, suggesting that sex differences in this variable are not the cause of burnout. The only antecedent of burnout to differ across sex was workload. Women in this study self-reported higher quantitative levels of workload ($d = -0.22$), and workload did relate to the individual facets of burnout, with the exception of personal efficacy. Though the observed differences in workload may be sample dependent, future researchers may benefit from investigating this relation.

The results of this study, although unsuccessful in their initial endeavor, are positive for the field of burnout research, as the findings of complete measurement invariance give confidence and support to researchers’ assessments of group mean differences in this area. Although two items from one of the burnout scales were found to exhibit statistically significant violations of invariance, the effect sizes for these items and the scale as a whole were nugatory. The other findings of this study also create a cohesive image of good psychometric properties of these scales. Specifically, no differential prediction was detected for any of the variables in this study that are in burnout’s nomological network. Thus, relations between burnout and its correlates are found to be valid across sex. Additionally, the use of a more complex IRT model (the Generalized Graded Unfolding Model) was not found to substantially improve model-data fit over a more traditional model (the Graded Response Model). This is important because,
though they do differ in other theoretical assumptions about item responses, classical test theory and traditional item response theory both agree that higher observed scores are monotonically positively reflective of higher levels of the latent construct. The inability of an unfolding response model to better represent the data gives further credence to the soundness of past and future findings in the area of burnout.

This study had several interesting and surprising findings in addition to the primary analyses. First, as evidenced by the analysis of 14 subject matter experts, the proportion of interpersonal content within individual burnout items has drastically reduced since the initial conceptualization of the construct. This may reflect the specific scales (e.g. the Maslach Burnout Inventory – General Survey as opposed to the Maslach Burnout Inventory – Human Services Survey) or it may reflect growing opinion in the field that burnout is not entirely what we once thought it was. Second, in support of the notion that burnout has evolved as a construct, one particular subscale, the Personal Efficacy subscale of the MBI, did not relate as strongly to the nomological network assessed as did the other facets of burnout across all three inventories. This supports the growing notion that personal efficacy is an outcome of burnout, not a core part of it. Lastly, interpersonal orientation was not related to any facet of burnout on any scale and failed to explain any variance in levels of burnout. This further supports burnout’s move away from its interpersonal roots.

Because it is clear from the results of this study that people of both sexes experience burnout, it is important to consider options for dealing with such cases when they arise. Awa, Plaumann, and Walter (2010) offer a review of various intervention programs designed to address burnout at the individual and organizational level. Their results showed that both levels of intervention can be effective at reducing burnout, though combined interventions (those
targeting specific employees with burnout as well as the organization as a whole) tend to have
longer lasting effects. Additionally, they recommend refresher courses to help prevent burnout,
as the effects of the interventions faded over time, regardless of the level of focus. Many of the
interventions reviewed by Awa, Plaumann, and Walter used the MBI to assess burnout; given the
recent trend away from the use of the MBI and its conceptualization of burnout, combined with
the results of this study that showed other scales with comparable psychometric properties, future
research and applied interventions should assess burnout using these alternative tools. The
review of interventions was also unable to assess whether intervention worked similarly for
people of both sexes. Given that this research corroborates extant sex differences in the
prevalence of burnout, such questions must be addressed. It is imperative, however, that men
and women receive equal opportunities for assessment and intervention.

Limitations and Future Research

This study did have several limitations that need to be addressed. First, the current
questionnaire was quite long (over 100 items) and the order of the scales was not randomized.
Although all burnout inventories were put at the beginning, some later scales may have been
affected by test-taker fatigue. On the same note, all participants received all burnout inventories
successively and in the same order, so there may have been some carry-over effects. Authors are
encouraged to replicate and expand the current research because, as noted above, even IRT
methods are subject to issues inherent in the sample. Second, the measure of interpersonal
orientation employed in this study had several contextual issues that suggested it is out of date,
with some items being egregious enough to warrant exclusion from the study. Supplementary
factor analyses of this scale yielded multiple factors, not just one as recommended by the original
authors. Additionally, the internal consistency of the scale was rather low considering the large
number of items in it, which may also be indicative of multiple factors being assessed. Further research and development of this scale is highly encouraged. Third, the sample sizes in this study are on the lower end of what is considered acceptable for IRT research, particularly for research using the Generalized Graded Unfolding Model, which typically needs upwards of 750 subjects for reliable estimates (Roberts, Laughlin, Donoghue, 2000). Lastly, over 80% of participants selected “Prefer not to respond” for the question asking about demographic; because of this, there is no way to determine exactly where the sample was primarily located, given that Mturk is open to people across the globe. As such, there may be unforeseeable cultural differences affecting results.

Although the three largest inventories of burnout were assessed here, there are still other measures of burnout that warrant investigation as well. Some, notably other versions of the MBI, may have higher levels of interpersonal content and may therefore have different results than those found here. Future research should work to expand measurement invariance assessments to these and other burnout scales to provide a more complete view of the psychometric properties of the research tools in this area. Additionally, it may be beneficial to control for job type by gathering participants from a single occupation to control for differences in the amounts of interpersonal stressors. This would give a stronger test of measurement invariance null hypotheses.

**Conclusion**

The results of this research, although failing to support the proposed hypotheses, provide a good outlook for burnout research as a whole. Researchers are encouraged to use the measures assessed here with greater confidence in the soundness of their psychometric properties while continuing to work at developing and refining these tools. As stated above, the results also do
not point to a clear cause of group mean differences in burnout but do corroborate their existence and, in light of the correlates assessed here, their importance. Further work investigating these differences and ways to combat them are strongly encouraged.


### Table 1

Means, standard deviations, and correlations of study variables

<table>
<thead>
<tr>
<th>Scale</th>
<th>M</th>
<th>SD</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>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CBI P</td>
<td>2.66</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></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 CBI W</td>
<td>2.64</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
<td></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>3 CBI C</td>
<td>2.50</td>
<td>0.93</td>
<td>.56**</td>
<td>.69**</td>
<td></td>
<td></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>4 OLBI D</td>
<td>2.41</td>
<td>0.57</td>
<td>.54**</td>
<td>.70**</td>
<td>.56**</td>
<td></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>5 OLBI EE</td>
<td>2.33</td>
<td>0.52</td>
<td>.67**</td>
<td>.82**</td>
<td>.60**</td>
<td>.76**</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>6 MBI EE</td>
<td>4.06</td>
<td>1.56</td>
<td>.66**</td>
<td>.82**</td>
<td>.58**</td>
<td>.71**</td>
<td>.82**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 MBI C</td>
<td>3.92</td>
<td>1.64</td>
<td>.53**</td>
<td>.67**</td>
<td>.56**</td>
<td>.76**</td>
<td>.70**</td>
<td>.76**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 MBI PE</td>
<td>2.40</td>
<td>1.06</td>
<td>.34**</td>
<td>.39**</td>
<td>.37**</td>
<td>.44**</td>
<td>.49**</td>
<td>.35**</td>
<td>.42**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 JDI</td>
<td>38.07</td>
<td>15.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.93</td>
</tr>
<tr>
<td>10 JIG</td>
<td>38.88</td>
<td>16.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.92</td>
</tr>
<tr>
<td>11 ITQ</td>
<td>3.31</td>
<td>1.91</td>
<td>.44**</td>
<td>.56**</td>
<td>.46**</td>
<td>.66**</td>
<td>.57**</td>
<td>.59**</td>
<td>.63**</td>
<td>.33**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 PSI</td>
<td>24.26</td>
<td>8.41</td>
<td>.60**</td>
<td>.61**</td>
<td>.46**</td>
<td>.48**</td>
<td>.56**</td>
<td>.58**</td>
<td>.47**</td>
<td>.28**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.89</td>
</tr>
<tr>
<td>13 ICAWS</td>
<td>7.44</td>
<td>3.08</td>
<td>.38**</td>
<td>.41**</td>
<td>.41**</td>
<td>.33**</td>
<td>.39**</td>
<td>.38**</td>
<td>.37**</td>
<td>.36**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.85</td>
</tr>
<tr>
<td>14 QWI</td>
<td>16.47</td>
<td>4.92</td>
<td>.33**</td>
<td>.41**</td>
<td>.29**</td>
<td>.28**</td>
<td>.40**</td>
<td>.39**</td>
<td>.29**</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.84</td>
</tr>
<tr>
<td>15 IOS</td>
<td>3.34</td>
<td>0.36</td>
<td>.06</td>
<td>.08</td>
<td>.01</td>
<td>-.03</td>
<td>.06</td>
<td>.05</td>
<td>-.03</td>
<td>-.10*</td>
<td>.20**</td>
<td>.04</td>
<td>.01</td>
<td>.04</td>
<td>-.02</td>
<td>.08</td>
<td>.69</td>
</tr>
</tbody>
</table>

Note: CBI = Copenhagen Burnout Inventory, CBI P = Personal Burnout, CBI W = Work-Related Burnout, CBI C = Client-Related Burnout; OLBI = Oldenburg Burnout Inventory, OLBI D = Disengagement, OLBI EE = Emotional Exhaustion; MBI = Maslach Burnout Inventory, MBI EE = Emotional Exhaustion, MBI C = Cynicism, MBI PE = Personal Efficacy; JDI = Job Descriptive Index (Coworker Subscale); JIG = Job in General Scale; ITQ = Intention to Quit Scale; PSI = Physical Symptoms Inventory; ICAWS = Interpersonal Conflict at Work Scale; QWI = Quantitative Workload Inventory; IOS = Interpersonal Orientations Scale; Values in parentheses are Cronbach’s alpha for each scale

*p < .05 **p < .01
Table 2
Significant group mean differences

<table>
<thead>
<tr>
<th>Scale</th>
<th>M(SD) males\textsuperscript{a}</th>
<th>M(SD) females\textsuperscript{b}</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBI P</td>
<td>2.57 (.69)</td>
<td>2.80 (.71)</td>
<td>-3.83</td>
<td>&lt;.001</td>
<td>-.32</td>
</tr>
<tr>
<td>CBI W</td>
<td>2.54 (.77)</td>
<td>2.79 (.90)</td>
<td>-3.52</td>
<td>&lt;.001</td>
<td>-.30</td>
</tr>
<tr>
<td>OLBI EE</td>
<td>2.27 (.49)</td>
<td>2.42 (.55)</td>
<td>-3.56</td>
<td>&lt;.001</td>
<td>-.30</td>
</tr>
<tr>
<td>MBI EE</td>
<td>3.94 (1.47)</td>
<td>4.24 (1.66)</td>
<td>-2.18</td>
<td>&lt;.001</td>
<td>-.19</td>
</tr>
<tr>
<td>PSI</td>
<td>23.33 (8.04)</td>
<td>25.66 (8.78)</td>
<td>-3.29</td>
<td>.001</td>
<td>-.28</td>
</tr>
<tr>
<td>QWI</td>
<td>16.04 (4.46)</td>
<td>17.13 (5.50)</td>
<td>-2.51</td>
<td>.012</td>
<td>-.22</td>
</tr>
<tr>
<td>IOS</td>
<td>3.29 (.36)</td>
<td>3.42 (.34)</td>
<td>-4.23</td>
<td>&lt;.001</td>
<td>-.36</td>
</tr>
</tbody>
</table>

\textsuperscript{a}n = 348 \quad \textsuperscript{b}n = 231
Table 3
Item-level fit for the GRM and GGUM

<table>
<thead>
<tr>
<th>Item</th>
<th>GRM</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
<td>$\chi^2$/df</td>
<td>$\chi^2$</td>
<td>df</td>
<td>$\chi^2$/df</td>
<td></td>
</tr>
<tr>
<td>CBI P 1</td>
<td>20.20</td>
<td>24</td>
<td>0.84</td>
<td>1.60</td>
<td>4</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20.40</td>
<td>26</td>
<td>0.78</td>
<td>0.93</td>
<td>4</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>36.72</td>
<td>33</td>
<td>1.11</td>
<td>0.10</td>
<td>4</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.41</td>
<td>29</td>
<td>0.70</td>
<td>0.32</td>
<td>4</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>20.64</td>
<td>23</td>
<td>0.90</td>
<td>0.53</td>
<td>4</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>49.70</td>
<td>35</td>
<td>1.42</td>
<td>0.12</td>
<td>4</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>CBI W 1</td>
<td>45.04</td>
<td>42</td>
<td>1.07</td>
<td>0.15</td>
<td>4</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>45.60</td>
<td>42</td>
<td>1.09</td>
<td>0.31</td>
<td>4</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>33.78</td>
<td>36</td>
<td>0.94</td>
<td>0.47</td>
<td>4</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>48.52</td>
<td>52</td>
<td>0.93</td>
<td>41.02</td>
<td>16</td>
<td>2.56</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>43.35</td>
<td>42</td>
<td>1.03</td>
<td>4.45</td>
<td>8</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>36.83</td>
<td>36</td>
<td>1.02</td>
<td>0.48</td>
<td>4</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>26.00</td>
<td>29</td>
<td>0.90</td>
<td>1.08</td>
<td>4</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>CBI C 1</td>
<td>35.16</td>
<td>27</td>
<td>1.30</td>
<td>1.25</td>
<td>4</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>38.75</td>
<td>24</td>
<td>1.61</td>
<td>1.00</td>
<td>4</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>23.57</td>
<td>23</td>
<td>1.02</td>
<td>1.66</td>
<td>4</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>58.04</td>
<td>44</td>
<td>1.32</td>
<td>24.69</td>
<td>16</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>31.94</td>
<td>31</td>
<td>1.03</td>
<td>0.23</td>
<td>4</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>49.38</td>
<td>38</td>
<td>1.30</td>
<td>0.20</td>
<td>4</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>OLBI EE 1</td>
<td>30.08</td>
<td>23</td>
<td>1.31</td>
<td>0.02</td>
<td>3</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>17.12</td>
<td>27</td>
<td>0.63</td>
<td>0.28</td>
<td>3</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>40.17</td>
<td>34</td>
<td>1.18</td>
<td>47.48</td>
<td>15</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.69</td>
<td>24</td>
<td>0.86</td>
<td>0.12</td>
<td>3</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>35.48</td>
<td>30</td>
<td>1.18</td>
<td>38.38</td>
<td>12</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>56.07</td>
<td>38</td>
<td>1.48</td>
<td>21.12</td>
<td>18</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>76.88</td>
<td>43</td>
<td>1.79</td>
<td>53.33</td>
<td>27</td>
<td>1.98</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>23.23</td>
<td>22</td>
<td>1.06</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>OLBI C 1</td>
<td>34.88</td>
<td>30</td>
<td>1.16</td>
<td>36.06</td>
<td>15</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>34.10</td>
<td>28</td>
<td>1.22</td>
<td>5.79</td>
<td>6</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>37.05</td>
<td>27</td>
<td>1.37</td>
<td>20.61</td>
<td>9</td>
<td>2.29</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>42.71</td>
<td>25</td>
<td>1.71</td>
<td>0.09</td>
<td>3</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>20.14</td>
<td>26</td>
<td>0.77</td>
<td>10.75</td>
<td>9</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>32.52</td>
<td>22</td>
<td>1.48</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>23.31</td>
<td>23</td>
<td>1.01</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>39.41</td>
<td>30</td>
<td>1.31</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>MBI EE 1</td>
<td>109.53</td>
<td>71</td>
<td>1.54</td>
<td>0.17</td>
<td>6</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>76.00</td>
<td>61</td>
<td>1.25</td>
<td>0.25</td>
<td>6</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>56.20</td>
<td>63</td>
<td>0.89</td>
<td>0.22</td>
<td>6</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>72.67</td>
<td>59</td>
<td>1.23</td>
<td>0.35</td>
<td>6</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>75.85</td>
<td>53</td>
<td>1.43</td>
<td>0.32</td>
<td>6</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>GRM $\chi^2$</td>
<td>df</td>
<td>$\chi^2/df$</td>
<td>GGUM $\chi^2$</td>
<td>df</td>
<td>$\chi^2/df$</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>----</td>
<td>------------</td>
<td>----------------</td>
<td>----</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>MBI C 1</td>
<td>96.49</td>
<td>64</td>
<td>1.51</td>
<td>1.64</td>
<td>6</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>98.02</td>
<td>68</td>
<td>1.44</td>
<td>1.69</td>
<td>6</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>165.37</td>
<td>91</td>
<td>1.82</td>
<td>148.34</td>
<td>48</td>
<td>3.09</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>90.97</td>
<td>75</td>
<td>1.21</td>
<td>121.46</td>
<td>48</td>
<td>2.53</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>105.02</td>
<td>80</td>
<td>1.31</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>MBI PE 1</td>
<td>84.43</td>
<td>51</td>
<td>1.66</td>
<td>1.03</td>
<td>6</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>87.87</td>
<td>63</td>
<td>1.39</td>
<td>42.00</td>
<td>24</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>80.61</td>
<td>48</td>
<td>1.68</td>
<td>0.81</td>
<td>6</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>113.03</td>
<td>79</td>
<td>1.43</td>
<td>125.46</td>
<td>66</td>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>98.51</td>
<td>68</td>
<td>1.45</td>
<td>28.34</td>
<td>18</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>106.87</td>
<td>47</td>
<td>2.27</td>
<td>1.54</td>
<td>6</td>
<td>0.26</td>
<td></td>
</tr>
</tbody>
</table>
### Table 4
Differential Item Functioning Statistics and Effect Sizes

<table>
<thead>
<tr>
<th>Item</th>
<th>Total $\chi^2$</th>
<th>df</th>
<th>$\alpha - \chi^2$</th>
<th>df</th>
<th>$\delta - \chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>Normed SIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBI C 3</td>
<td>15.9</td>
<td>7</td>
<td>0.1</td>
<td>1</td>
<td>0.78</td>
<td>6</td>
<td>.03</td>
<td>.021</td>
</tr>
<tr>
<td>MBI C 4</td>
<td>16.6</td>
<td>7</td>
<td>6.6</td>
<td>1</td>
<td>10.0</td>
<td>6</td>
<td>.01</td>
<td>.009</td>
</tr>
</tbody>
</table>
Figure 1
Item characteristic curve with DIF for item 3 of the MBI Cynicism subscale

Note: Dashed lines indicate trace lines for females. Solid lines indicate trace lines for males.
Figure 2
Item characteristic curve with DIF for item 4 of the MBI Cynicism subscale

Note: Dashed lines indicate trace lines for females. Solid lines indicate trace lines for males.
Figure 3
Expected score curve for item 1 of the CBI Work-Related Burnout Subscale

Item Characteristic Curve

Item#: 01, delta= 3.432, alpha= 1.501, tau= 0.000, -5.337, -4.041, -2.631, -1.205.
Figure 4
Category response probability curve for item 1 of the CBI Work-Related Burnout Subscale under the Graded Response Model
Figure 5
Category response probability curve for item 1 of the CBI Work-Related Burnout Subscale under the Generalized Graded Unfolding Model
APPENDIX B

Copenhagen Burnout Inventory

*Personal burnout (α=.83)*
- How often do you feel tired? a
- How often are you physically exhausted? a
- How often are you emotionally exhausted? a
- How often do you think: “I can’t take it anymore”? a
- How often do you feel worn out? a
- How often do you feel weak and susceptible to illness? a

*Work-related burnout (α=.88)*
- Do you feel worn out at the end of the working day? a
- Are you exhausted in the morning at the thought of another day at work? a
- Do you feel that every working hour is tiring for you? a
- Do you have enough energy for family and friends during leisure time? a (R)
- Is your work emotionally exhausting? b
- Does your work frustrate you? b
- Do you feel burnt out because of your work? b

*Client-related burnout (α=.91)*
- Do you find it hard to work with clients? b
- Does it drain your energy to work with clients? b
- Do you find it frustrating to work with clients? b
- Do you feel that you give more than you get back when you work with clients? b
- Are you tired of working with clients? a
- Do you sometimes wonder how long you will be able to continue working with clients? a

<table>
<thead>
<tr>
<th>Response categories and scoring:</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Never/Almost Never</td>
</tr>
<tr>
<td>Often</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sometimes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seldom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To a very high degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To a high degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somewhat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To a low degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To a very low degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Oldenburg Burnout Inventory

I always find new and interesting aspects in my work. ¹
There are days when I feel tired before I arrive at work. (R) ²
It happens more and more often that I talk about my work in a negative way. (R) ¹
After work, I tend to need more time than in the past in order to relax and feel better. (R) ²
I can tolerate the pressure of my work very well. ²
Lately, I tend to think less at work and do my job almost mechanically. (R) ¹
I find my work to be a positive challenge. ¹
During my work, I often feel emotionally drained. (R) ²
Over time, one can become disconnected from this type of work. (R) ¹
After working, I have enough energy for my leisure activities. ²
Sometimes I feel sickened by my work tasks. (R) ¹
After my work, I usually feel worn out and weary. (R) ²
This is the only type of work that I can imagine myself doing. ¹
Usually, I can manage the amount of my work well. ²
I feel more and more engaged in my work. ¹
When I work, I usually feel energized. ²

1 indicates disengagement items (α=.82)
2 indicates exhaustion items (α=.84)

<table>
<thead>
<tr>
<th>Response categories and scoring:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>Strongly Agree</td>
</tr>
<tr>
<td>Agree</td>
</tr>
<tr>
<td>Disagree</td>
</tr>
<tr>
<td>Strongly Disagree</td>
</tr>
</tbody>
</table>
Job Descriptive Index
(Note: All subscales ask the respondent if the word does or does not describe the facet in question)

People on Your Present Job (α=.93)

<table>
<thead>
<tr>
<th>Stimulating</th>
<th>Likeable</th>
<th>Unpleasant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boring</td>
<td>Intelligent</td>
<td>Supportive</td>
</tr>
<tr>
<td>Slow</td>
<td>Easy to make enemies</td>
<td>Active</td>
</tr>
<tr>
<td>Helpful</td>
<td>Rude</td>
<td>Narrow interests</td>
</tr>
<tr>
<td>Stupid</td>
<td>Smart</td>
<td>Frustrating</td>
</tr>
<tr>
<td>Responsible</td>
<td>Lazy</td>
<td>Stubborn</td>
</tr>
</tbody>
</table>

Job in General (α=.92)

<table>
<thead>
<tr>
<th>Pleasant</th>
<th>Worthwhile</th>
<th>Makes me content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td>Worse than most</td>
<td>Inadequate</td>
</tr>
<tr>
<td>Great</td>
<td>Acceptable</td>
<td>Excellent</td>
</tr>
<tr>
<td>Waste of time</td>
<td>Superior</td>
<td>Rotten</td>
</tr>
<tr>
<td>Good</td>
<td>Better than most</td>
<td>Enjoyable</td>
</tr>
<tr>
<td>Undesirable</td>
<td>Disagreeable</td>
<td>Poor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response categories and scoring:</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
</tr>
<tr>
<td>Y</td>
</tr>
</tbody>
</table>
Intention to Quit ($\alpha = .97$)

I intend to leave this organization soon.

I plan to leave this organization in the next little while.

I will quit this organization as soon as possible.

I do not plan on leaving this organization soon. (R; italics in original)

I may leave this organization before too long.

<table>
<thead>
<tr>
<th>Response categories and scoring:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Mildly Disagree</td>
<td>Neither Agree nor Disagree</td>
<td>Mildly Agree</td>
<td>Agree</td>
<td>Strongly Agree</td>
<td></td>
</tr>
</tbody>
</table>
Physical Symptoms Inventory – 13-item Version ($\alpha=.89$)

Over the past month, how often have you experienced each of the following?

An upset stomach or nausea
A backache
Trouble sleeping
Headache
Acid indigestion or heartburn
Eye strain
Diarrhea
Stomach cramps (Not menstrual)
Constipation
Ringing in the ears
Loss of appetite
Dizziness
Tiredness or fatigue

<table>
<thead>
<tr>
<th>Response Categories and Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
</tbody>
</table>
Interpersonal Conflict at Work Scale (α=.85)

How often do you get into arguments with others at work?
How often do other people yell at you at work?
How often are people rude to you at work?
How often do other people do nasty things to you at work?

<table>
<thead>
<tr>
<th>Response Categories and Scoring:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>Rarely</td>
<td>Sometimes</td>
<td>Quite Often</td>
<td>Very Often</td>
<td></td>
</tr>
</tbody>
</table>
Quantitative Workload Inventory, QWI ($\alpha=0.84$)

How often does your job require you to work very fast?
How often does your job require you to work very hard?
How often does your job leave you with little time to get things done?
How often is there a great deal to be done?
How often do you have to do more work than you can do well?

<table>
<thead>
<tr>
<th>Response Categories and Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Less than once per month or never</td>
</tr>
</tbody>
</table>
Interpersonal Orientations Scale ($\alpha=.69$) (1-5, *Strongly Disagree to Strongly Agree*)

I would rather think about a personal problem by myself than discuss it with others. (R)

I consider myself a forgiving person.

I am interested in knowing what makes people tick.

When I receive a gift, I find myself thinking about how much it must be worth. (R)

Under no circumstances would I buy something I suspected had been stolen.

I am greatly influenced by the moods of the people I am with.

Sometimes the most considerate thing one person can do for another is to hide a bit of the truth.

Sometimes simply talking aloud about things that bother me makes me feel better – regardless of who, if anyone, hears these thoughts.

My friends and I seem to share the same musical interests.

I am reluctant to talk about my personal life with people I do not know well. (R)

I generally view myself as a person who is not terribly interested in what other people are really like. (R)

I sometimes think I take things that other people say to me too personally.

It’s important for me to work with people with whom I get along well, even if that means I get less done.

If I were to share an apartment with somebody, I would want to find out about the person’s family background, hobbies, and so forth.

I tend to like people who are good looking.

What others think about my actions is of little or no consequence to me. (R)

The more other people reveal about themselves, the more inclined I feel to reveal things about myself.

When someone does me a favor I don’t usually feel compelled to return it. (R)

Sitting on a bus or a subway, I sometimes imagine what the person sitting next to me does for a living.

The more I am with others, the more I tend to like them.

I would rather be given a simple and thoughtful gift than a more extravagant one that involved less thought and care.

I am very sensitive to criticism.

When people tell me personal things about themselves, I find myself feeling close to them.

One good turn does not necessarily deserve another. (R)

I can be strongly affected by someone smiling or frowning at me.