AN ITEM RESPONSE THEORY ANALYSIS OF CWB MEASUREMENT ARTIFACTS

Stacy Sim

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

December 2016

Committee:

Clare Barratt, Advisor

Michael Zickar

William O'Brien
ABSTRACT

Clare Barratt, Advisor

Researchers have increasingly raised concerns about the measurement of counterproductive work behavior (CWB; e.g., Bowling & Gruys, 2010). In particular, one issue that has yet to be examined is the effect of measurement properties on CWB responding. This issue warrants immediate attention because researchers often modify aspects of the measure without empirically confirming their validity or equivalence. Common modifications include the number of response options or the referent time frame in the instructions. Thus, the objective of the current study was to apply item response theory (IRT) to a 2 x 3 experimental design in order to psychometrically examine the causal effects of (a) a 5-point versus 7-point Likert scale and (b) referent time frames of employment length, one year, or one month on a CWB measure. Classical Test Theory and IRT analyses were conducted. Results indicated that modifications to these measurement properties affected responses at the item level, as well as correlations with a common correlate. Overall, the modified measures assessed lower levels of the CWB latent trait than the original measure, which required a much higher level of the latent trait in order for respondents to endorse response options. Implications for research and practice are discussed.
ACKNOWLEDGMENTS

I would like to express my deepest gratitude and appreciation for all my mentors, friends, and family members, who have selflessly and continuously invested in my growth and success over the years. I attribute all my successes to the combination of their influences on my life.

Beginning chronologically, I have the most amazing family to thank first and foremost for shaping me into the person that I am today. My parents, Kiam Hwee and Suzanne, have always role modeled what dedication and hard work looks like, with the singular motivation of love and the associated desire to create a better life for their two children. Through their unconditional love and support, I have had the invaluable freedom to dream and pursue any path I dared to envision before me. Along the same lines, I have all the appreciation in the world for my sister, Melissa, for being my biggest cheerleader and always providing practical advice and optimism to improve any situation I find myself in.

Next, my undergraduate mentor, Dr. Frank Bernieri, has my utmost respect and gratitude for instilling in me a passion for research, teaching me to understand the world through the fascinating lens of social psychology, and providing me with tremendous opportunities to better myself. I am certain I would not be where I am without his investment and confidence in me.

I would also like to thank BGSU and all my faculty and peers for this exceptional opportunity to earn my doctorate degree while gaining the experience, knowledge, and skills to pursue my passion and be a credit to my field. My thesis advisor, Dr. Clare Barratt, deserves my highest gratitude for taking me on as her first student and being incredibly supportive of all my professional pursuits. Her time and guidance have been instrumental in me slowly but surely finding my place in the field of I/O psychology. Lastly, to Scott and all my wonderful friends who have seen me through my trials, thank you for extending me your friendship, grounding me, laughing with me, and making this journey truly enjoyable. I hope to make you all proud.
# 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>Popular CWB Measures</td>
<td>2</td>
</tr>
<tr>
<td>Workplace Deviance Scale</td>
<td>3</td>
</tr>
<tr>
<td>Counterproductive Work Behavior-Checklist</td>
<td>4</td>
</tr>
<tr>
<td>Measurement Artifacts</td>
<td>5</td>
</tr>
<tr>
<td>Response Scale</td>
<td>5</td>
</tr>
<tr>
<td>Referent Time Frame</td>
<td>6</td>
</tr>
<tr>
<td>Item Response Theory (IRT)</td>
<td>8</td>
</tr>
<tr>
<td>Summary</td>
<td>10</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>10</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>10</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>11</td>
</tr>
<tr>
<td>Research Question 4</td>
<td>11</td>
</tr>
<tr>
<td>METHOD</td>
<td>12</td>
</tr>
<tr>
<td>Design</td>
<td>12</td>
</tr>
<tr>
<td>Participants and Procedure</td>
<td>12</td>
</tr>
<tr>
<td>Measures</td>
<td>13</td>
</tr>
<tr>
<td>Counterproductive Work Behavior-Checklist</td>
<td>13</td>
</tr>
<tr>
<td>Counterproductive Work Behavior-Checklist, Modified Versions</td>
<td>13</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td>14</td>
</tr>
<tr>
<td>Demographics</td>
<td>14</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>RESULTS</td>
<td>15</td>
</tr>
<tr>
<td>Correlational Analyses</td>
<td>15</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>15</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>15</td>
</tr>
<tr>
<td>Item Response Theory</td>
<td>15</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>16</td>
</tr>
<tr>
<td>Research Question 4</td>
<td>18</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>21</td>
</tr>
<tr>
<td>Limitations</td>
<td>25</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>28</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>30</td>
</tr>
<tr>
<td>APPENDIX A. ORIGINAL COUNTERPRODUCTIVE WORK BEHAVIOR-CHECKLIST</td>
<td>51</td>
</tr>
<tr>
<td>APPENDIX B. JOB SATISFACTION</td>
<td>53</td>
</tr>
<tr>
<td>APPENDIX C. CONSENT FORM</td>
<td>54</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                    Page
1  Manipulated Conditions of the Study ................................................................. 38
2  Measurement Artifacts Manipulations................................................................. 39
3  Descriptive Statistics of All Variables by Condition ......................................... 40
4  Comparisons of Mean CWB by Condition ............................................................. 41
5  Correlations Between CWB and Job Satisfaction by Condition .............................. 42
6  Average Item Parameters for Response Scale and Referent Time Frame Conditions ................................................................. 43
7  Item Information Frequencies for Response Scale and Referent Time Frame Conditions ................................................................. 44
8  Test Information Functions for Response Scale and Referent Time Frame Conditions Across the Latent Trait Continuum ................................................................. 45
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Examples of Low and High Information</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>Item Information Function for Response Scale</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>Test Information Functions Across Response Scale</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>Item Information Function for Referent Time Frame</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>Test Information Functions Across Referent Time Frame</td>
<td>50</td>
</tr>
</tbody>
</table>
INTRODUCTION

Counterproductive work behavior (CWB) refers to volitional acts that intend to harm or cause actual harm to an organization and/or its stakeholders (Spector & Fox, 2005). CWBs can take many different forms and can be directed at other employees or the organization. Researchers and practitioners study CWB because its outcomes are overwhelmingly detrimental for both individuals and organizations. These outcomes include overt anger (Aquino, Douglas, & Martinko, 2004), anxiety and depression (Duffy, Ganster, Shaw, Johnson, & Pagon, 2006; Lim & Lee 2011), decreased self-esteem, self-efficacy, organizational commitment, and increased turnover intentions (Duffy, Ganster, & Pagon, 2002; Houshmand, O’Reilly, Robinson, & Wolff, 2012; Leiter & Maslach, 1988; Low, Radhakrishnan, Schneider, & Rounds, 2007). As a result of CWB, Fox and Stallworth (2004) reported that 77% of people dreaded going in to work, 74.8% seriously considered quitting their job, and 66% got headaches, upset stomach, or chest pains on the job over the past five years. Further, these CWBs have been estimated to cost organizations $4.2 billion in lost productivity and CWB-related legal expenses (Bensimon, 1994), and affect employees’ reputations and morale (Filipczak, 1993). Due to the excessive direct costs to employee wellbeing and organizational overhead, researchers and practitioners have focused their attention and efforts on predicting perpetrators of CWB and understanding the motivations behind engaging in it (e.g., Mount, Ilies, & Johnson, 2006; Spector et al., 2006).

Despite the increase in research on the topic of CWB, researchers continue to raise concerns over the measurement of its construct. In general, these measurement studies have focused on identifying potential moderators of CWB effects. Examples include (a) relevance of the behavior to the job (Bowling & Gruys, 2010), (b) use of self-report (Berry, Carpenter, & Barratt, 2012), and (c) rating type (Bennett & Robinson, 2000; Dalal, 2005; Gruys, 1999; Gruys
& Sackett, 2003; Spector et al., 2006). However, to date, CWB investigations have not examined the effect of inherent psychometric properties, which is the proposed foundation upon which further investigations are built upon. Although identifying moderators is empirically valuable to reconcile discrepancies in the literature and understand the factors that affect relationships, differences in the measure itself should be explored first to ensure that properties of the measures are not producing artifactual covariance in observed relationships (Podsakoff et al., 2003). Thus, the objective of the current study is to conduct a preliminary exploration of the effects of measurement properties on CWB responding in an experimental design from a psychometric approach using item response theory (IRT).

**Popular CWB Measures**

The exploration of measurement artifacts on differences in CWB effect sizes should be framed within the discussion of how CWB is measured. In general, researchers have commonly administered one of two CWB measures to assess self-reports of employee CWB: the Workplace Deviance Scale (WDS; Bennett & Robinson, 2000) and the Counterproductive Work Behavior Checklist (CWB-C; Spector et al., 2006). The WDS (Bennett & Robinson, 2000) was published in 2000 and quickly rose in popularity due to its intuitive conceptualization of CWB. However, this soon gave way to the latter measure, which is the main measure that dominates the CWB literature today.

Even though both measures purport to measure CWB, there is elementary evidence to support the contention that the two measures are not functionally equivalent. Although the focal point of Carpenter and Berry’s (2014) meta-analysis was to distinguish between CWB and withdrawal, they found that choice of scale was a moderator, where the CWB-withdrawal effect size was higher when measured with the CWB-C (ρ = .68) than with the WDS (ρ = .45). This
finding begs the questions of *why* two measures of CWB result in such different effect sizes, and which version is more accurate. Immediately concluding that the two CWB measures are incompatible on the basis of this finding would be premature since it is possible that there were systematic differences in the measurement properties of the scale or the measures themselves. Thus, the current study will provide a closer examination of differences in the two CWB scales to determine whether measurement properties of the scales contribute to differential effect sizes.

**Workplace Deviance Scale.** Robinson and Bennett (1995) initially applied a multidimensional scaling technique and proposed a model of CWB in which the behaviors fell on two orthogonal dimensions: (a) target (interpersonal versus organizational), and (b) severity (minor versus severe). However, they later refined their model because they noted that severity reflected a quantitative description of CWB as opposed to a distinctive category (Bennett & Robinson, 2000). Thus, their creation of the WDS in 2000 distinguished between CWB that targeted other employees and those that targeted the organization. This categorization was intuitive and easy for researchers and practitioners to conceptualize, which contributed to its general acceptance and popular administration.

The WDS consists of 19 items, 7 of which reflect behaviors targeted at individuals in the organization, and 12 reflect behaviors that target the organization itself. Researchers have reported factor analytic and differential correlational support with other variables for this two-factor distinction (Bennett & Robinson, 2000; Berry, Ones, & Sackett, 2007; Gruys & Sackett, 2003; Robinson & Bennett, 1995). Respondents are instructed to rate the frequency with which they engaged in each behavior in the **past year** using a **7-point** Likert scale with benchmarks of 1 (never), 2 (once a year), 3 (twice a year), 4 (several times a year), 5 (monthly), 6 (weekly), and 7 (daily).
**Counterproductive Work Behavior-Checklist.** The CWB-C was created to address concerns that the two-factor model would be insufficient in its simplicity to accurately reflect true relationships with other antecedents and outcomes. For example, they expressed concern about the low base rates of CWB when the two target-focused factors were broken down into finer subdivisions based on severity (i.e., severe versus minor). Furthermore, Spector et al. (2006) posited that employees were more likely to demonstrate consistency in their behavior and engage in the same type of CWB (e.g., theft), instead of performing a wide range of behaviors that appear largely dissimilar. As such, they created the CWB-C, which recategorized CWB into five behavioral distinctions based on the type of CWB: abuse, withdrawal, production deviance, sabotage, and theft. Since its publication, the CWB-C has quickly risen in popularity. In part, this may be because it allowed researchers and practitioners to make finer predictions of specific CWB, which are useful in creating interventions targeted at specific types of CWB, such as theft (Sackett & DeVore, 2001).

The CWB-C (Spector et al., 2006; Spector, Bauer, & Fox, 2010) has three different versions - the 45-item checklist (CWB-C-45), the 32-item checklist (CWB-C-32), and a shortened 10-item checklist (CWB-C-10). Whereas the CWB-C-45 conforms to the two-factor, target-focused conceptualization of CWB (CWB-I versus CWB-O), the CWB-C-32 utilizes a five-factor model that categorizes items according to its type of CWB - abuse, withdrawal, production deviance, sabotage, and theft. Support for the two and five factor model has been suggested by differential correlations with other variables (e.g., job satisfaction, conflict, justice, emotions; Spector et al., 2006). Given the following prompt, “How often have you done each of the following things on your present job?”, respondents are instructed to rate each item on a 5-point Likert scale with benchmarks of 1 (never), 2 (once or twice), 3 (once or twice per month),
4 (once or twice per week), and 5 (every day). The CWB-C-45 will be used in this study in order to control for the difference in conceptualization (i.e., target versus behavior) of CWB and enable a parallel comparison with the WDS.

Although the WDS and CWB-C have several similarities, their main differences that will be studied experimentally in the current investigation are the effects of (a) number of response options and (b) referent time frame in the instructions, on how employees respond to CWB items. Since researchers frequently modify these properties in their studies without determining their equivalence (e.g., Aquino, Galperin, & Bennett, 2004; Chandler, 2008; Cohen, Panter, & Turan, 2013; Dunlop & Lee, 2004), this investigation will inform researchers on the impact measurement properties may have on results. In extension, findings contribute the potential to assuage discrepancies among existing studies using different measures of CWB. Moreover, this will equip researchers to make appropriate decisions about altering properties of their measure.

**Measurement Artifacts**

**Response Scale.** The first measurement artifact that is frequently modified is the response scale, which refers to the number of response options respondents are presented with. Commonly used response scales are the 5-point and 7-point Likert scales, which correspond to the CWB-C (Spector et al., 2006) and WDS (Bennett & Robinson, 2000) respectively. Although some researchers have found little effect for the use of different response scales (e.g., Bendig, 1953, 1954; Brown, Wilding, & Coulter, 1991), others have acknowledged that the number of response options can influence psychometric properties (e.g., Dawes, 2008; Lissetz and Green, 1975; Preston & Colman, 2000). In general, researchers have reported that the internal consistency, test-retest reliability, and convergent and criterion validity tend to increase with more response options, however there is a lack of consensus as to the location of the asymptote,
as well as what the optimal number of response options are (e.g., Chang, 1994; Hancock & Klockars, 1991; Preston & Colman, 2000). For example, Dawes (2008) examined the effect of response scale on data characteristics in a price consciousness scale, chosen for its easily understandable content for respondents. He found no difference between the 5-point and 7-point scales in terms of mean scores (after being rescaled for comparison), variance, skewness, or kurtosis, but reported that there were slight differences with a 10-point scale. Preston & Colman (2000), on the other hand, found that several indices (e.g., reliability, validity, discriminating power, test-retest reliability) declined after 7 response options, although internal consistency was comparable across all response options. In part, these discrepancies across studies have been explained by the usage of different criteria, (e.g., internal consistency, test-retest reliability, convergent validity, user experience), methods (computer simulation, hypothetical experiences, real life experiences), and variables (Preston & Colman, 2000), which demonstrate that there is no single, optimal number of response options that generalizes across all constructs and situations. Thus, the examination of the effect of response scale on CWB responding is warranted for the potential effect that response scale may have on the psychometric properties of the scale and relationships with other variables. The current study will examine the psychometric effects of the two most commonly used response options, 5- versus 7-point Likert scale, on patterns of responses at the item and scale level using item response theory.

Referent Time Frame. The second measurement artifact that is frequently modified is the referent time frame, which instructs respondents to relay their engagement in CWB for a specified time period. The current study will examine the effect of referent time frames of employment length, one-year, and one-month. As with the previous measurement artifact, the former referent time frames (i.e., employment length and one-year) are based on the two
commonly administered measures – the WDS (Bennett & Robinson, 2000) and the CWB-C (Spector et al., 2006). However, researchers often take liberties in utilizing items from the original CWB measure but applying an alternative referent time frame. Recently, there has been an increasing shift towards using a referent time frame of one-month out of concern for the transient nature of memory, as well as for the feasibility for longitudinal designs (e.g., Matthews, Kath, & Barnes-Farrell, 2010; Sakurai & Jex, 2012; Tepper & Henle, 2011). Since longitudinal studies require at least three data points, referent time frames of one-year and one’s entire employment length severely undermine, if not eliminate, the possibility of collecting repeated data points. Thus, it is of practical importance to evaluate the impact of utilizing a referent time frame of one-month, instead of longer traditional time frames of employment length and one-year.

According to Podsakoff et al. (2003), selecting an ideal referent time frame has important implications for CWB research because it serves as a cue that will determine what events are retrieved from memory. In general, shorter reference periods that are clearly specified tend to produce more accurate reports of behavior than longer periods (e.g., Del Boca & Darkes, 2003; Morin, 1993; Neath, 1993). However, this seems to be true only of common behaviors that follow a typical pattern (Bradburn et al., 1987; Menon, 1997; Sudman & Bradburn, 1982). In contrast, for more atypical behaviors, like CWB, a short recall period may be more subject to error if there is a mismatch between the referent time frame and the typicality of the period for that behavior (Timotijevic, Barnett, Shephard, & Senior, 2009). For example, self-reports of common behaviors like “daydreaming while at work” might be more accurately captured if measured on a daily basis, whereas less frequent behaviors, such as “blaming someone for an error you made”, might not be adequately captured if the measurement period (e.g., daily diary
over a week) precludes the behavior. Further, although each individual may have their own ideal referent time frame, overall, some time frames may be more appropriate than others. For CWB specifically, an ideal time frame should be long enough to allow events that are representative of the job to have occurred, yet short enough that memory effects would not substantially bias CWB reporting. Time frames that are inappropriately long might be too vague for respondents to recall specific events, whereas ones that are too short might be insufficient to representatively reflect general CWB tendencies. In either situation, choosing an inappropriate referent time frame would limit the accuracy of CWB results (e.g., Timotijevic et al., 2009). However, to date, researchers have yet to empirically investigate the extent to which referent time frames affect participant responses to CWB measures and items. Therefore, the current study will address this gap in the literature using IRT as the framework through which actual response patterns can be examined for its impact.

**Item Response Theory (IRT)**

Item response theory (IRT) is a modern psychometric approach that uniquely “establish[es] the correspondence between latent variables and their manifestations” (de Ayala, 2009, p. 4) at an item level. It is commonly applied to understand how people respond to items based on their levels on a latent variable, which is invaluable for developing and scoring tests. In the present study, IRT will be applied to understand how people are responding to CWB items. Specifically, the item response functions (IRF), item information functions (IIF), and test information functions (TIF) will be examined to explore how measurement properties affect CWB responding.

The item response function (IRF) refers to the relationship between an underlying latent trait (denoted by $\theta$) and the probability of endorsing a given response (bounded by 0 to 1), which
are plotted on the x-axis and y-axis respectively. An IRF is generated for each item, and the location of the curves can be visually inspected to examine the pattern of CWB responding across response options. For example, if response options group on the left side of the figure, this will indicate that a low level of the latent trait is needed for endorsement, such that even individuals with low scores on the latent trait have a high probability of selecting the response option. In contrast, response options that group on the right side of the figure will indicate that a high level of the latent trait is needed for endorsement. In other words, respondents need to have high levels of the latent trait before they will endorse that response option. Furthermore, the flatness or steepness of a curve, which is a function of discrimination \( (a) \), will indicate the ability of that item to differentiate between individuals with different levels of the latent trait, whereas the item location parameters \( (b) \) reflect the inflection point of the curve, which corresponds to the threshold point used to determine the likelihood of endorsement of a response option by a respondent of a given latent trait level.

Based on the premise that not all items are created equally, some items will provide greater precision in their estimate of item parameters (e.g., \( \theta \)) than others (Baker, 2001). Items that have lower variability in their estimates of the value of the parameter are said to provide more information. To determine whether the level of information provided is potentially affected by measurement artifacts at the item and test level, the item information functions (IIF) and test information functions (TIF) will be analyzed. The IIF provides a graph of information as a function of the latent trait. This can be used to identify the amount of information that is provided for different levels of the latent trait, which can be useful for inferring whether the item is unequally targeting subpopulations of respondents with certain levels of the latent trait. In comparison, TIF is calculated as an aggregate of the IIF, which is important to examine because
items on a test contain different amounts of information at varying levels across the latent continuum. These interpretations are influenced by the discrimination \((a)\) and item location \((b)\) parameters, such that higher discrimination parameters reflect more information at a narrower range, and the item location parameter determines the level of the latent trait that can be estimated with the highest precision.

**Summary**

The objective of the current study is to experimentally examine the effect of measurement properties on responses to a CWB measure and its items. Specifically, this study will address the causal influences of (a) response scale (5-point versus 7-point Likert scale), and (b) referent time frame in the instructions (employment length, one-year, one-month) on CWB’s relationships with other variables and responses to CWB items.

First, correlations between CWB and a common correlate (i.e., job satisfaction) will be investigated in order to estimate the effect of response scale and referent time frame modifications on the current effect sizes in the literature. Specifically, job satisfaction was identified as a variable of interest due to its practical relevance to organizations and employees. Furthermore, meta-analytic correlations of the CWB-job satisfaction relationship (Mean \(\rho = -.37\); e.g., Dalal, 2005) can be compared to the effect sizes of the current study to provide an estimate of how skewed existing results might be as a result of measurement artifacts.

**Research Question 1.** How do response scales (5- versus 7-point Likert scale) affect the relationship between CWB-C and job satisfaction?

**Research Question 2.** How do referent time frames (1 month, 1 year, employment length) affect the relationship between CWB-C and job satisfaction?
Next, IRT analyses will be conducted to explore the effects of measurement properties on CWB responding at an item and test level. Specifically, the item response functions (IRF), item information functions (IIF), and test information functions (TIF) will be examined to determine whether modified CWB measures affect the levels of information provided across comparable ranges of CWB.

**Research Question 3.** How do response scales (5- versus 7-point Likert scale) affect responses to CWB items via (a) item response functions (IRF), (b) item information functions (IIF), and (c) test information functions (TIF)?

**Research Question 4.** How do referent time frames (1 month, 1 year, employment length) affect responses to CWB items via (a) item response functions (IRF), (b) item information functions (IIF), and (c) test information functions (TIF)?
METHOD

Design

This study utilized a 2 (response scale) x 3 (referent time frame) between-subjects experimental design to causally determine the effect of measurement artifacts on CWB responding. Based on the measurement properties of the two most commonly administered measures of CWB, this study examined the effects of response scales (5 and 7-point Likert scales) and referent time frames (overall employment length, one-year, one-month) on relationships with job satisfaction and response patterns at the item and test level.

Participants and Procedure

Participants were 1,569 Amazon Mechanical Turk (MTurk) respondents who were screened for at least part-time employment status in an organization and located in the United States of America. There were 841 men (56%) and 650 women (43%); six participants refrained from selecting an option. Their mean age was 35.33 years old (SD = 10.60) and the majority of participants were Caucasian (77%). They were employed at the organization for 5.51 years on average (SD = 4.00), with the largest proportion of them working 31-40 hours a week (50.7%).

Data were collected in multiple waves in order to reduce sampling error and maximize the representativeness of the sample since data are collected fairly quickly on MTurk. Participants were randomly assigned to one of six possible versions of the CWB measure, which was important for the purpose of the current study (i.e., to evaluate the equivalence of CWB responses and scores across modified versions; e.g., Holland, Dorans, & Peterson, 2007). Participants were administered the measures via an online Qualtrics survey, which was hosted on MTurk. The survey took less than 5 minutes to complete and $1.00 was offered in exchange for their participation. Seventy-two participants were screened out for having zero variance across
measures since it likely reflected careless responding. The final sample consisted of 1,497 participants with 244 to 255 participants in each condition.

Measures

**Counterproductive Work Behavior-Checklist.** The CWB-C was the original two-factor measure containing 45 items (Spector et al., 2006; See Appendix A). The original measure utilizes a referent time frame of respondents’ length of employment. According to the instruction, “How often have you done each of the following things on your present job?”, respondents rate each item on a 5-point scale. The current study utilized benchmarks of 1 (never), 3 (occasionally), and 5 (everyday) in order to enable equitable comparisons (i.e., control for additional measurement artifacts) across manipulated conditions (discussed below).

Researchers have reported high internal consistency for the CWB-C (e.g., Fox, Spector, Goh, Bruursema & Kessler, 2012), and the Cronbach alpha was .97 in the current study.

**Counterproductive Work Behavior-Checklist, Modified Versions.** The second version (CWB-C-V2) and third version (CWB-C-V3) of the CWB-C maintained the original CWB-C, 5-point response scale, but had different referent time frames of one-year and one-month, respectively. In contrast, the fourth (CWB-C-V4), fifth (CWB-C-V5), and sixth (CWB-C-V6) versions of the CWB-C employed a 7-point rating scale but referenced the respective time frames of employment length, one-year, and one-month. Table 1 provides a summary of these design manipulations, and Table 2 details the specific modifications performed on the response scale and referent time frame. The CWB-C (and its alternate versions) demonstrated high item reliability in the current sample. The overall Cronbach alpha was .97 (range of .96-.98 over the modified conditions).
**Job Satisfaction.** Following Judge and colleagues’ example, a 5-item measure of Brayfield and Rothe’s (1951) job satisfaction scale was used (see Appendix B; e.g., Judge & Ilies, 2004; Judge, Parker, Colbert, Heller, & Ilies, 2001). Participants rated each item on a 5-point scale from strongly disagree to strongly agree. A sample item was “I find real enjoyment in my work”. This measure demonstrated high item reliability in the current sample, where the overall Cronbach alpha was .92 (range of .91-.93 across the six conditions).

**Demographics.** In addition to the above measures, demographic variables such as age, sex, race, employment status, and tenure were also collected.
RESULTS

Table 3 reports the descriptive statistics for all measures and Table 4 summarizes the conditions whose means significantly differed from others based on a one-way ANOVA.

Correlational Analyses

Research Question 1. The aim of the first research question was to explore the effect of response scale on the relationship between CWB and job satisfaction. Table 5 reports individual correlations for each of the six conditions. However, for the purpose of the current research question, correlational analyses were conducted after collapsing across the three referent-time frame conditions in order to compare the effect sizes for a 5-point scale versus a 7-point scale. Based on a significance test of effect sizes, the mean correlation between CWB and job satisfaction for the 5-point scale \( r = -.24 \) did not significantly differ from the 7-point scale \( r = -.26; z = -1.87, p > .05 \).

Research Question 2. The second research question sought to explore the effect of referent time frame on the relationship between CWB and job satisfaction. Again, correlations were conducted after collapsing across response scale in order to determine the CWB-job satisfaction relationships for one-month, one-year, and employment length referent time frames. The mean correlation between CWB and job satisfaction for the one-year \( r = -.20 \) and employment length \( r = -.21 \) did not differ from one another \( z = -.29, p > .05 \). However, the one-month \( r = -.30 \) effect size differed significantly from the one-year \( r = -.20 \) and employment length \( r = -.21 \) conditions \( z = 2.75, p < .01, \) and \( z = 2.45, p < .05 \), respectively.

Item Response Theory

First, to test the unidimensionality assumption of IRT for appropriate model selection, factor analyses were conducted. For both the overall sample \( N = 1,497 \) and within each
condition (N’s of 244-255), the first eigenvalue accounted for more than 20% of the variance, and there was a substantial decrease in variance explained by subsequent eigenvalues (i.e., no more than 2-3% for the second eigenvalue; e.g., Scherbaum, Cohen-Charash, & Kern, 2006). Thus, the unidimensionality of CWB was supported and IRT analyses could be conducted to examine the research questions.

Since IRT items suffer from indeterminacy, meaning that scaling constants are arbitrary (Holland et al., 2007), item parameters were equated in IRTPRO 3 (Cai, Thissen, & du Toit, 2011) before analyses were conducted to ensure that values could be comparable across the six conditions. Since CWB measures are measured on a Likert scale and do not have response choices that are equidistant from one another, the graded response model (GRM; Samejima, 1969) was selected since it is appropriate for data with response choices that are polytomous (e.g., Likert scale) and not equidistant from one another. Since the GRM was applied, each item will have multiple threshold parameters.

Research Question 3. The purpose of the third research question was to explore the effects of response scale on CWB items and the overall measure using IRT. The GRM applied to the data exhibited an excellent fit overall, based on a $\chi^2$/df ratio of less than 2.0 (Drasgow & Hulin, 1990). Out of the 45 items, only two items from the 5-point scale analysis and four items from the 7-point scale analysis had $\chi^2$/df ratios that were greater than 2 (but less than 2.6). Since the $\chi^2$/df ratios for all items were less than 3, the GRM fit was well within the recommended threshold so analyses could be pursued.

Table 6 provides the average discrimination parameters across all items, as well as the average threshold parameters. Both average discrimination parameters for the 5-point and 7-point scales were on the high end of the normal discrimination range, meaning that on average,
the items in both conditions were able to discriminate well between various levels of the latent trait. However, there were small differences between the 5- and 7-point scales, such that the average discrimination parameter was marginally higher for the 5-point scale \((a = 2.6)\) than the 7-point scale \((a = 2.4)\). To approximate item locations, the thresholds of the GRM were averaged. The average item location parameters for the 7-point scale ranged from -1.5 to 1.6, whereas the range was 0.7 to 3.2 for the 5-point scale. Thus, the 7-point scale covered a greater range of the latent trait than the 5-point scale (range = 3.1 vs. 2.5).

Information is a continuous function, where higher values are more desirable. However, to compare the item information functions across conditions, information was categorized as low and high based on the number of items with maximum information lower than 2 and greater than 3 respectively (See Table 7). Visually, items with a maximum information amount of less than 2 tend to appear homogeneously flat across the latent continuum and were not considered to contribute a substantial amount of information. In contrast, items with maximum peaks higher than 3 were considered to contribute a meaningful amount of information about the construct (although maximum information levels higher than 2 were also recognized as an alternative threshold; See Figure 1). Although both the 5-point and 7-point scales had 10 items that each had information higher than 3, the 5-point scale had fewer items \((k = 20)\) that had information lower than 2 than the 7-point scale \((k = 25\) items). Figure 2 provides a visual comparison of the response scales using item 16 ("Failed to report a problem so it would get worse") as an example, where the latent trait continuum is plotted on the x-axis and the information levels provided by that item. This figure demonstrates that the 5-point scale captures information across a higher and narrower range of the underlying latent trait, whereas the 7-point scale provided
information over a lower to moderate, and broader range of CWB. This general pattern was replicated across multiple items that did not have flat item information functions.

Item information functions can be aggregated across items to provide an information function for the entire test (i.e., test information function; See Table 8). The test information function figures follow the same format as that of the item information functions, such that the latent trait continuum is plotted on the x-axis and the information levels provided by the test is plotted on the y-axis. Figure 3 contains the test information functions for the 5- and 7-point scales, demonstrating that on a test level, the aggregated information across the 45 items had a higher maximum for the 5-point scale (information = 101) than the 7-point scale (information = 92). These peaks also corresponded to different latent trait levels, such that the 5-point scale maximum occurred at the latent trait level of 2, and the 7-point scale maximum occurred at the latent trait level of -0.4. In general, since tests with more information are better able to measure the construct, higher information levels are a desirable characteristic to the extent that they correspond to the desirable range of the latent trait that is of interest to the researcher.

**Research Question 4.** The last research question sought to examine the effects of referent time frame on CWB items and the overall measure. Holistically, the GRM fit the data well. With the exception of two items (i.e., item 22, “Took supplies or tools home without permission” and item 24, “Put in to be paid for more hours than you worked”) that were removed from the analysis for poor fit (i.e., $\chi^2$/df ratio greater than 3.0), the remaining 43 items had a $\chi^2$/df ratio of less than 3.0, with the majority having a ratio less than 2.0, indicating an excellent fit (Drasgow & Hulin, 1990; Drasgow, Levine, Tsien, Williams, & Mead, 1995).

As in the last research question, the discrimination parameters were averaged across all items (Table 6). Similar to the results for response scales, the average discrimination parameters
for the referent time frames (i.e., employment length, one-year, and one-month) were on the high end of the normal discrimination range, indicating that the items discriminate well between various levels of the latent trait. Marginal differences between the referent time frames existed, where the average discrimination parameter across the items was highest for the employment length time \((a = 2.9)\), followed by the one-month \((a = 2.7)\), and then the one-year \((a = 2.3)\), describing the average amount of information the items have over a narrow range, in descending order. To approximate item locations, the thresholds of the GRM were averaged. The average item location parameter \((b)\) ranged from 0.5 to 3.5 for the employment length time (range of \(b = 3.0\)), -1.5 to 1.9 for the one-year (range of \(b = 3.4\)), and -1.4 to 1.4 for the one-month (range of \(b = 2.8\)). These parameters indicate that the employment length time frame covered moderate to very high levels of the latent trait, the one-year condition covered low to high levels of the latent trait, and the one-month condition covered low to moderate levels of the latent trait.

To compare the item information functions, the number of items with maximum information lower than 2 and higher than 3 were again compared (See Table 7). Whereas both the employment length and one-month times had an equal number of items with information higher than 3 \((k = 17)\) and lower than 2 \((k = 20)\), the one-year referent time had only 4 items with information higher than 3 and 27 that had information lower than 2. Figure 4 provides a visual representation of the amount of information \((y\text{-axis})\) provided across the latent continuum \((x\text{-axis})\) by the three referent time frames using item 15 (“Purposely came late to an appointment or meeting”) as an example. Whereas the one-year and one-month referent time frames had similar patterns of information at a lower to moderate range of CWB, the employment length referent time frame had a higher amount of information over higher latent trait levels.
To assess the amount of information at the test level, the item information functions were aggregated across all items (See Table 8). Following the same format as the item information functions, Figure 5 depicts the amount of information provided by the test (y-axis) across the latent continuum (x-axis). At the test level, the referent time frame of one’s entire employment had the highest maximum information of 125, followed by the one-month (information = 108) and one-year (information = 78). These levels corresponded to different latent trait levels, at 1.6, -0.4, and -0.4 respectively. Although the one-month and one-year had maximum information at the same latent trait level and a comparable pattern of information at the test level, the one-month had slightly higher information than the one-year condition.
DISCUSSION

Despite the increasing amount of research conducted on CWB, many measurement issues have yet to be empirically examined and resolved. Given the common application of modifications to scale characteristics, such as response scale and referent time frame, the objective of the present study was to explore the impact of measurement properties on CWB results and apply IRT to examine patterns of responses to CWB items and the overall measure. Separate analyses were conducted to determine the effect of response scale (data collapsed across referent time frame) and the effect of referent time frame (data collapsed across response scale).

First, correlational analyses were conducted to assess the impact of modifying measurement properties on CWB relationships. Whereas response scale (5-point vs. 7-point) did not cause differential effect sizes between CWB and job satisfaction (i.e., the correlation was very similar for the 5-point scale and the 7-point scale), the second measurement property of referent time frame did affect the magnitude of effect sizes. Specifically, the CWB-job satisfaction correlation for the one-month referent time frame was significantly higher than both the employment length and one-year referent time frames. This was consistent with the overall examination of the six conditions (response scale by referent time frame), as conditions that utilized a month referent time frame had CWB-job satisfaction correlations that were significantly higher than the other two time frames.

On one hand, these results may raise potential implications for current methodological initiatives that encourage researchers to sample shorter time intervals to increase the feasibility of longitudinal designs, since relationships may appear stronger than if they were assessed using referent time frames of respondents’ employment length or one-year, which are the default time
frames of the original, two most commonly administered CWB measures (i.e., CWB-C and WDS). However, on the other hand, compared to the meta-analytic correlation between CWB and job satisfaction of -.37 (Dalal, 2005), the overall mean correlation of -.23 in the current sample across all the conditions was substantially lower, with the correlation from the one-month referent time frame conditions of -.30 bearing the greatest resemblance to the meta-analytic correlation. These results could be interpreted in several ways. One possibility is that the one-month referent time is a beneficial modification that most closely approximates meta-analytic effect sizes. This would be a favorable outcome since this shorter time length enables the implementation of longitudinal designs. Another possibility, however, is that the meta-analytic correlation most closely resembles the one-month referent time frame because the studies it consisted of had many modifications to one-month referent time frames, and meta-analyses are descriptive of the population to the extent that individual studies are representative of the phenomenon of interest. If this is the case, based on the pattern of effect sizes in the current study, one might expect the meta-analytic correlation to be lower than the one-month effect size since it is composed of a multiple referent time frames. However, this might also be a function of this specific sample and the specific job satisfaction measure used in this study, as opposed to the variety of measures used across studies in a meta-analysis. Nevertheless, this hypothesis can be empirically tested by conducting a meta-analysis using the same, if not more studies, and examining referent time frame as a moderator.

To examine the effect of measurement properties on item responding, IRT analyses were conducted. Although the average item discrimination parameters were comparably high for the 5- and 7-point scales, the average item location parameter ($b$) of the 5-point scale covered a narrower range of the latent trait than the 7-point scale, and exclusively covered the positive
range of the latent trait. In practical terms, the decision to use a 5- or 7-point Likert scale should depend on the specificity of the latent trait levels of the desired sample. For example, if the research question involves assessing respondents with a wider range of latent traits, these findings suggest that a 7-point scale might be more appropriate since it would be similarly discriminating as a 5-point scale, but would enable the researchers to more precisely differentiate between lower to moderate levels of the latent trait.

Based on the item information functions, the 5-point scale had a greater number of items with a maximum information level higher than 3 than the 7-point scale. Consistently, at the test level, the 5-point scale had a higher information peak than the 7-point scale, although they differed in terms of the corresponding level of the latent trait. Whereas the information maximum occurred at a latent trait level of 2.0 for the 5-point scale, it corresponded to the latent trait level of -0.4 for the 7-point scale. The 5-point scale provided more information than the 7-point scale at its peak (i.e., less measurement error), which occurred at a higher level of the latent trait than the 7-point scale. Again, depending on the latent trait level of interest, researchers should be aware that changing the response scale can impact the amount of information obtainable across the latent continuum even though it did not result in differential relationships with job satisfaction in this case.

In a similar pattern, although the discrimination parameters for the employment length, one-year, and one-month referent time frames fell on the high end of the normal discrimination range, the employment length condition had the highest average discrimination (a), followed by the one-month and the one-year referent time frames. Contrasting the estimated item location parameters based on the averaged thresholds demonstrated that the one-year condition exhibited the greatest range in average thresholds, followed by the employment length, and one-month
referent time frames. In general, having a wider range may be desirable because such a test would enable you to assess people of more varied levels of CWB. However, where the item locations are along the latent trait also matter, as they reflect differences in the latent trait levels that can be more precisely measured. In this case, the employment length time frame covered moderate to very high levels of the latent trait, the one-year condition covered low to high levels of the latent trait, and the one-month condition covered low to moderate levels of the latent trait. Depending on the research question of interest (e.g., high performers of CWB or average performers of CWB), modifications to the referent time frame can facilitate more precise estimates of parameters.

The employment length and one-month referent time frames had the same number of items with high information, which was greater than the one-year condition. However, even though the frequencies of the items with high information was the same between the employment length and one-month referent time frames, the maximum information was the highest for the employment length, followed by the one-month, and one-year conditions. Whereas the location of the peak information along the latent trait continuum was the same for the one-year and one-month at -0.4, the employment length condition had maximum information at a latent trait level of 1.6. Even though the employment length time frame had the highest maximum information, it corresponded to a moderately high level of the latent trait, whereas the one-month time frame had the next highest maximum information and corresponded to a moderately low level of the latent trait. Again, depending on whether the research question surrounds extreme performers of CWB or average performers of CWB, researchers should be aware of the effect of these modifications on the scale so that they can make informed decisions in order to maximize the precision with which these parameters can be estimated. Although modifications to the response
scale did not cause differential relationships with job satisfaction, changes to the referent time frame did. The CWB-job satisfaction correlation was significantly higher than both the employment length and one-year referent time frames, indicating that the scale characteristic of referent time frame had an effect on not only CWB responding, but also resultant relationships with other constructs.

**Limitations**

Despite the contributions of this paper to understanding the causal effect of measurement properties on responses to a CWB scale at the item and test level, this study was not without its limitations. First, the sample consisted of MTurk workers due to the large number of participants needed for estimation accuracy. This presented several issues – given the short survey times and low incentives for accurate responding, careless responding was a concern for data quality. Attempts to address this consisted of additional screening during data cleaning to minimize careless responding. For example, participants who had zero variance across both the CWB and job satisfaction measures were inferred as likely indicating click-through responding. Despite the small possibility that these responses truly reflected a few participants’ behaviors and attitudes, it is likely that the number of observed zero variance cases exceeded the number of true zero variance cases, so a decision rule was created to remove all participants who had zero variance across both measures in order to be safe and minimize error. Nevertheless, these efforts to identify and remove careless responding may not have been entirely successful – although internal consistency is expected to increase with additional items, the Cronbach alpha for the CWB scale was still extremely high at .97, which may be an indicator of click-through responding. Furthermore, irrespective of the accuracy of responses, the data may be specific to this MTurk sample. Although it is possible that the results may be unrepresentatively biased
towards individuals who have low CWB and are therefore willing to answer these questions or individuals with personality disorders who are not ashamed of their behaviors, efforts were taken to ensure participants of the anonymity of their responses to minimize sample-specific concerns.

Secondly, response scales were displayed using the anchors of ‘never’ and ‘daily’, with a middle benchmark of ‘occasionally’. This adds a component of subjectivity for respondents to individually determine what the response options in between the benchmarks provided are. Hakel (1968) reported a surprising amount of ambiguity in respondents’ perceptions of commonly used frequency words. Although they seemed to agree upon the frequency of extreme anchors such as “always” and “never”, there was much less consensus for other words (e.g., sometimes, occasionally, frequently, about as often as not, etc.), which were perceived to occur at very different frequencies. However, this design was optimal for the current study since using the traditional CWB-C benchmarks would be confounded with the referent time frame, and the objective of this study was to investigate the causal effects of these two measurement properties independent from one another. If future studies confirm the greater influence of referent time frame than response scale on CWB results, future studies can increase experimental control by examining a single measurement property at a time to minimize the introduction of further errors.

Findings from this study can be applied to inform researchers and practitioners on the effects of making seemingly small modifications to their measures. For example, if the intent is to identify employees with high CWB tendencies or measure more extreme CWB behaviors, then using a scale with characteristics (i.e., 5-point scale and employment length time frame) that capture higher levels of the latent trait would be beneficial. On the other hand, if the question of interest involves lower performers of CWB, then utilizing other scale characteristics would be
more beneficial. Ultimately, the underlying rational would be to increase the utility of a measure by maximizing the alignment of scale characteristics with the desired latent trait levels of the construct of interest.
CONCLUSION

The purpose of this study was to explore the effects of response scale and referent time frame on CWB responding. Although there were differences in the item response functions and item and test information functions between the 5-point and 7-point response scales, these differences were not reflected in the correlations with job satisfaction. However, this was not the case for the referent time frame, as there were differences between referent time conditions in both the IRT analyses and the correlational analyses. Whereas there was no difference between the employment length and one-year effect sizes, the correlations based on the one-month referent time frame were significantly higher than that of the employment length and one-year referent time frames.

The comprehensive examination of both correlational analyses and IRT approaches was beneficial to not only guide future research, but also to retrospectively understand how modifications to the CWB-C scale may affect our understanding of relationships reported in the literature. It is important for researchers to not only be aware of the potential effects of measurement properties on effect sizes and responses, but to also take precautions to verify the equivalence of their modified version with the original measure, or perhaps even demonstrate an improvement of their modified measure over the original measure for their specific purposes. Future research is needed to understand the mechanism behind the effect of measurement properties on results and the extent to which it affects current and future studies. For example, a meta-analysis could be conducted to investigate the potential moderator of referent time frame to determine the extent to which this property affects current meta-analytic correlations. Alternatively, another experimental study can be conducted with more controls in order to focus
on the specific effect of response scale on CWB responses and relationships using a different sample.

Overall, the contribution of applying IRT to understand how participants are responding is helpful to not only shorten measures by eliminating items with low information, but also to understand what measurement properties are most desirable for the range of the latent trait of interest. Understanding how items work at various levels of the latent trait continuum is beneficial since it likely sets the boundaries for which meaningful data can be obtained. Thus, even though IRT is likely resource-intensive in terms of expertise, sample size, data collection costs, and time, it is also a valuable approach that would greatly improve the measurement of psychological constructs and effectively further the field of Psychology.
REFERENCES


Chandler, M. M. (2008). Examining the mechanisms by which situational and individual difference variables relate to workplace deviance: the mediating role of goal self-concordance. The University of Akron. Retrieved from
https://etd.ohiolink.edu!etd.send_file?accession=akron1216846236&disposition=attachment


Dawes, J. G. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research, 50*(1), 61-77.


Table 1. Manipulated Conditions of the Study

<table>
<thead>
<tr>
<th>Referent time frame</th>
<th>Response scale</th>
<th>5-point</th>
<th>7-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment length</td>
<td>CWB-C</td>
<td>CWB-C-V4</td>
<td></td>
</tr>
<tr>
<td>One year</td>
<td>CWB-C-V2</td>
<td>CWB-C-V5</td>
<td></td>
</tr>
<tr>
<td>One month</td>
<td>CWB-C-V3</td>
<td>CWB-C-V6</td>
<td></td>
</tr>
</tbody>
</table>

Note: CWB-C = counterproductive work behavior checklist, original. The labels of V2-V6 denote variations of the scale. Each row represents a different referent time frame manipulation, whereas every column represents a response scale manipulation.
### Table 2. Measurement Artifacts Manipulations

<table>
<thead>
<tr>
<th>Measurement artifact</th>
<th>Manipulations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response scale</strong></td>
<td>Never</td>
</tr>
<tr>
<td>5-point</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Never</td>
</tr>
<tr>
<td>7-point</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Referent time frame</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment time</td>
<td>How often have you done each of the following things on your <em>present job</em>?</td>
</tr>
<tr>
<td>One-year</td>
<td>How often have you done each of the following things in the <em>past year</em>?</td>
</tr>
<tr>
<td>One-month</td>
<td>How often have you done each of the following things in the <em>past month</em>?</td>
</tr>
</tbody>
</table>
Table 3. Descriptive Statistics of All Variables by Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>CWB</th>
<th>Job satisfaction</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$M(SD)$</td>
<td>Cronbach $\alpha$</td>
<td>$N$</td>
<td>$M(SD)$</td>
<td>Cronbach $\alpha$</td>
</tr>
<tr>
<td>5-point, E.L.</td>
<td>252</td>
<td>1.66(.61)</td>
<td>.97</td>
<td>252</td>
<td>3.41(.99)</td>
<td>.93</td>
</tr>
<tr>
<td>5-point, 1-year</td>
<td>250</td>
<td>1.72(.71)</td>
<td>.97</td>
<td>250</td>
<td>3.39(.97)</td>
<td>.91</td>
</tr>
<tr>
<td>5-point, 1-month</td>
<td>255</td>
<td>1.59(.61)</td>
<td>.97</td>
<td>255</td>
<td>3.35(.98)</td>
<td>.92</td>
</tr>
<tr>
<td>7-point, E.L.</td>
<td>246</td>
<td>1.96(.90)</td>
<td>.97</td>
<td>246</td>
<td>3.53(.96)</td>
<td>.92</td>
</tr>
<tr>
<td>7-point, 1 year</td>
<td>244</td>
<td>2.14(1.03)</td>
<td>.98</td>
<td>244</td>
<td>3.41(.99)</td>
<td>.93</td>
</tr>
<tr>
<td>7-point, 1-month</td>
<td>250</td>
<td>1.89(.91)</td>
<td>.96</td>
<td>250</td>
<td>3.49(.98)</td>
<td>.92</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td><strong>1,497</strong></td>
<td><strong>1.82(.83)</strong></td>
<td><strong>.97</strong></td>
<td><strong>1,497</strong></td>
<td><strong>3.43(.98)</strong></td>
<td><strong>.92</strong></td>
</tr>
</tbody>
</table>

*Note. CWB = counterproductive work behavior. E.L. = Employment Length*
<table>
<thead>
<tr>
<th></th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 5-point, E.L.</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. 5-point, 1-year</td>
<td>-.06 (.07)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. 5-point, 1-month</td>
<td>.07 (.07)</td>
<td>.13 (.07)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. 7-point, E.L.</td>
<td>-.29 (.07)**</td>
<td>-.24 (.07)*</td>
<td>-.36 (.07)**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5. 7-point, 1 year</td>
<td>-.47 (.07)**</td>
<td>-.42 (.07)**</td>
<td>-.54 (.07)**</td>
<td>-.18 (.07)</td>
<td>-</td>
</tr>
<tr>
<td>6. 7-point, 1-month</td>
<td>-.23 (.07)*</td>
<td>-.17 (.07)</td>
<td>-.30 (.07)**</td>
<td>.06 (.07)</td>
<td>.25 (.07)*</td>
</tr>
</tbody>
</table>

*Note.* E.L. = Employment length.

Positive numbers reflect a higher mean in the row condition number compared to the column condition number.

*p < .05. **p < .01. ***p < .001.*
Table 5. Correlations between CWB and Job Satisfaction by Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-point, E.L.</td>
<td>252</td>
<td>-.23***</td>
</tr>
<tr>
<td>5-point, 1-year</td>
<td>250</td>
<td>-.18**</td>
</tr>
<tr>
<td>5-point, 1-month</td>
<td>255</td>
<td>-.32***</td>
</tr>
<tr>
<td>7-point, E.L.</td>
<td>246</td>
<td>-.23***</td>
</tr>
<tr>
<td>7-point, 1 year</td>
<td>244</td>
<td>-.22***</td>
</tr>
<tr>
<td>7-point, 1-month</td>
<td>250</td>
<td>-.32***</td>
</tr>
<tr>
<td>Overall</td>
<td>1,497</td>
<td>-.23***</td>
</tr>
</tbody>
</table>

*Note. CWB = counterproductive work behavior. E.L. = Employment length

*p < .05. **p < .01. ***p < .001.
Table 6. Average Item Parameters for Response Scale and Referent Time Frame Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
<td>$b_1$</td>
<td>$b_2$</td>
<td>$b_3$</td>
<td>$b_4$</td>
<td>$b_5$</td>
<td>$b_6$</td>
</tr>
<tr>
<td><strong>Response Scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-point</td>
<td>2.6</td>
<td>0.7</td>
<td>1.5</td>
<td>2.3</td>
<td>3.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7-point</td>
<td>2.4</td>
<td>-1.5</td>
<td>-0.8</td>
<td>-0.4</td>
<td>0.3</td>
<td>1.0</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Referent Time Frame</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment Length</td>
<td>2.9</td>
<td>0.5</td>
<td>1.2</td>
<td>1.7</td>
<td>2.5</td>
<td>3.2</td>
<td>3.5</td>
</tr>
<tr>
<td>One-year</td>
<td>2.3</td>
<td>-1.5</td>
<td>-0.8</td>
<td>-0.2</td>
<td>0.5</td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td>One-month</td>
<td>2.7</td>
<td>-1.4</td>
<td>-0.8</td>
<td>-0.3</td>
<td>0.4</td>
<td>1.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

*Note.* The number of $b$’s correspond to the number of response scale options minus 1. Item parameters reported in this table are represented as an average across all items.
Table 7. Item Information Frequencies for Response Scale and Referent Time Frame Conditions

<table>
<thead>
<tr>
<th>Information Level</th>
<th>Response Scale</th>
<th>Referent Time Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5-point</td>
<td>7-point</td>
</tr>
<tr>
<td>&gt; 3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>&gt; 2</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>&lt; 2</td>
<td>20</td>
<td>25</td>
</tr>
</tbody>
</table>

*Note.* Information lower than 2 was considered low, whereas information higher than 2 or 3 was considered high (See Figure 1).
Table 8. Test Information Functions for Response Scale and Referent Time Frame Conditions Across the Latent Trait Continuum

<table>
<thead>
<tr>
<th></th>
<th>0:</th>
<th>-2.8</th>
<th>-2.4</th>
<th>-2.0</th>
<th>-1.6</th>
<th>-1.2</th>
<th>-0.8</th>
<th>-0.4</th>
<th>0.0</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
<th>2.8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response Scale</strong></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>
<td></td>
</tr>
<tr>
<td>5-point</td>
<td>1.6</td>
<td>2.0</td>
<td>2.6</td>
<td>3.6</td>
<td>5.3</td>
<td>8.0</td>
<td>12.8</td>
<td>22.0</td>
<td>39.3</td>
<td>65.3</td>
<td>89.4</td>
<td>100.7</td>
<td>101.3</td>
<td>97.8</td>
<td>93.7</td>
<td></td>
</tr>
<tr>
<td>7-point</td>
<td>9.5</td>
<td>15.4</td>
<td>26.1</td>
<td>43.3</td>
<td>65.4</td>
<td>84.7</td>
<td>91.6</td>
<td>91.1</td>
<td>89.8</td>
<td>87.8</td>
<td>81.1</td>
<td>61.6</td>
<td>39.4</td>
<td>22.6</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td><strong>Referent Time Frame</strong></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>
<td></td>
</tr>
<tr>
<td>E.L.</td>
<td>1.7</td>
<td>2.2</td>
<td>3.0</td>
<td>4.3</td>
<td>6.3</td>
<td>9.7</td>
<td>16.2</td>
<td>30.4</td>
<td>60.0</td>
<td>100.3</td>
<td>122.4</td>
<td>124.6</td>
<td>120.1</td>
<td>116.0</td>
<td>107.5</td>
<td></td>
</tr>
<tr>
<td>One-year</td>
<td>10.9</td>
<td>17.3</td>
<td>27.9</td>
<td>43.3</td>
<td>60.6</td>
<td>73.5</td>
<td>77.7</td>
<td>77.1</td>
<td>75.0</td>
<td>74.2</td>
<td>68.7</td>
<td>55.6</td>
<td>37.9</td>
<td>23.1</td>
<td>13.6</td>
<td></td>
</tr>
<tr>
<td>One-month</td>
<td>8.3</td>
<td>14.0</td>
<td>26.1</td>
<td>49.9</td>
<td>81.9</td>
<td>103.1</td>
<td>107.6</td>
<td>103.6</td>
<td>104.1</td>
<td>96.9</td>
<td>66.0</td>
<td>36.8</td>
<td>20.1</td>
<td>11.2</td>
<td>6.5</td>
<td></td>
</tr>
</tbody>
</table>

*Note. E.L. = Employment Length*
Figure 1. Examples of Low and High Information

Note. CWB4 (on the left; i.e., “Told people outside the job what a lousy place you work for”) was considered low information (information level lower than 2), whereas CWB2 (in the middle; i.e., “Daydreamed rather than did your work”) and CWB19 (on the right; i.e., “Left work earlier than you were allowed to”) represent two thresholds for high information (information higher than 2 and 3, respectively).
Figure 2. Item Information Function for Response Scale

Note. The 7-point scale is the thicker line across the range of latent trait continuum, whereas the 5-point scale is the thinner line on the right.

CWB16 = “Failed to report a problem so it would get worse.”
Figure 3. Test Information Functions Across Response Scale

Note. Group 1 = 5-point scale. Group 2 = 7-point scale.
Figure 4. Item Information Function For Referent Time Frame

Note. The employment length referent time frame is the thinner line at the upper end of the latent trait continuum, whereas the one-year and one-month time frames are the thicker lines across the range of latent trait continuum. The line with slightly more information is the one-month time frame, whereas the smoother bimodal function with lower information is the one-year condition.

CWB15 = “Purposely came late to an appointment or meeting.”
Figure 5. Test Information Functions Across Referent Time Frame

Note. Group 1 = Employment length. Group 2 = One-year. Group 3 = One-month.
APPENDIX A. ORIGINAL COUNTERPRODUCTIVE WORK BEHAVIOR-CHECKLIST

How often have you done each of the following things on your present job? Use the following scale in your ratings (Spector et al., 2006):

Rating scale:
1 = never
2 = once or twice
3 = once or twice per month
4 = once or twice per week
5 = everyday

1. Purposely wasted your employer’s materials/supplies.
2. Daydreamed rather than did your work.
3. Complained about insignificant things at work.
4. Told people outside the job what a lousy place you work for.
5. Purposely did your work incorrectly.
6. Came to work late without permission.
7. Stayed home from work and said you were sick when you weren’t.
8. Purposely damaged a piece of equipment or property.
9. Purposely dirtied or littered your place of work.
10. Stolen something belonging to your employer.
11. Started or continued a damaging or harmful rumor at work.
12. Been nasty or rude to a client or customer.
13. Purposely worked slowly when things needed to get done.
14. Refused to take on an assignment when asked.
15. Purposely came late to an appointment or meeting.
16. Failed to report a problem so it would get worse.
17. Taken a longer break than you were allowed to take.
18. Purposely failed to follow instructions.
19. Left work earlier than you were allowed to.
20. Insulted someone about their job performance.
22. Took supplies or tools home without permission.
23. Tried to look busy while doing nothing.
24. Put in to be paid for more hours than you worked.
25. Took money from your employer without permission.
26. Ignored someone at work.
27. Refused to help someone at work.
28. Withheld needed information from someone at work.
29. Purposely interfered with someone at work doing his/her job.
30. Blamed someone at work for error you made.
31. Started an argument with someone at work.
32. Stole something belonging to someone at work.
33. Verbally abused someone at work.
34. Made an obscene gesture (the finger) to someone at work.
35. Threatened someone at work with violence.
36. Threatened someone at work, but not physically.
37. Said something obscene to someone at work to make them feel bad.
38. Hid something so someone at work couldn’t find it.
39. Did something to make someone at work look bad.
40. Played a mean prank to embarrass someone at work.
41. Destroyed property belonging to someone at work.
42. Looked at someone at work’s private mail/property without permission.
43. Hit or pushed someone at work.
44. Insulted or made fun of someone at work.
45. Avoided returning a phone call to someone you should at work.
APPENDIX B. JOB SATISFACTION

Instructions (Modified from Brayfield & Rothe, 1951): Some jobs are more interesting and satisfying than others. We want to know how people feel about different jobs. This section contains five statements about jobs. Use a five-point scale (strongly disagree to strongly agree) to describe how you feel about your present job. There are no right or wrong answers. We should like your honest opinion on each one of the statements.

1=strongly disagree
2=disagree
3=undecided
4=agree
5=strongly agree

1. I feel fairly satisfied with my present job.
2. Most days I am enthusiastic about my work.
3. Each day at work seems like it will never end.
4. I find real enjoyment in my work.
5. I consider my job to be rather unpleasant.
APPENDIX C. CONSENT FORM

Department of Psychology
Bowling Green State University
Bowling Green, OH 43403

Informed Consent

The purpose of this study is to apply modern statistical techniques to investigate the measurement properties of scales. This study is being conducted by Stacy Sim, a graduate student in the psychology department at Bowling Green State University, and overseen by her advisor, Dr. Clare Barratt.

It is important to examine whether properties of a scale matter because they may affect the accuracy of researchers’ findings. However, measures are often created without studying the effect of these properties. Thus, the goal of this research is to apply modern statistical techniques to examine how the properties of a measure affect each item to help researchers create better measures in the future.

As a participant, you will be asked to complete a short survey that contains three measures and a few demographic questions. The survey should take about 5 minutes to complete, and you will receive $1.00 for your participation.

You must be 18 years or older to participate in this study and work at least part-time for an organization. The risks associated with participating in this study are no more than would be encountered in every day life. Your participation in this study is completely voluntary, and you are free to discontinue participation in this study at any time. Deciding to participate or not will not affect any relationship you may have with Bowling Green State University. You may also freely decline to respond to any questions. Completing the survey indicates your consent to participate in this study.

To protect your anonymity, your data will be stored on password-protected computers of the researchers involved in the project. For your security, after you finish the survey, it is recommended that you clear your browser history and page cache to ensure that your information will not be stored. In addition, it is recommended that you complete the survey using a personal computer to minimize the potential that an employer or coworker would gain access to your responses on the chance that they have tracking software installed in their company equipment.

If you have any questions or comments regarding this study or your participation in it, you may contact the principal investigator, Stacy Sim, at sims@bgsu.edu or at 419-372-2301, or the principal investigator’s advisor, Dr. Clare Barratt at cbarrat@bgsu.edu or (419) 372-4250. If you have any questions about the conduct of this study or your rights as a research participant, you may contact the Chair of Bowling Green State University’s Human Subjects Review Board at 419-372-7716 or hrsb@bgsu.edu. Thank you for your time.

By clicking "next," you are consenting to participate in this study.