

**FACTORS AFFECTING SELF-OTHER AGREEMENT ABOUT EMPLOYEES'  
COUNTERPRODUCTIVE WORK BEHAVIOR: AN ITEM-LEVEL FOCUS**

A Thesis

by

JUAN CARLOS BATARSE

Submitted to the Office of Graduate and Professional Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,  
Co-chair of Committee,  
Committee Members,  
Head of Department,

Kathi N. Miner  
Christopher M. Berry  
Deidra J. Schleicher  
Douglas W. Woods

May 2015

Major Subject: Psychology

Copyright 2015 Juan Carlos Batarse

## **ABSTRACT**

The purpose of this study is to extend the literature on measurement of counterproductive work behavior (CWB), particularly the relationship between ratings of CWB from self- and other-rating (i.e., supervisors and coworkers) sources. Previous research suggests that self-ratings and other-ratings of CWB are usually viable alternatives to each other, but that they do contain sizable unique information. However, it remains unclear exactly what this unique information is and whether it is valid unique information. Therefore, there is a need to determine what are the differences and similarities in information contained in self- and other-ratings of CWB. The goal of the present study is to address this by examining the factors that affect relationships and mean differences between self-ratings and other-ratings of CWB at the item-level. The present study proposes 10 dimensions (e.g., observability, memorability, task relevance) along which CWB items can be categorized and that may affect self-other agreement. In the first study, subject matter experts rated each CWB item in the Bennett and Robinson CWB scale on the 10 dimensions. In the second study, 85 paired self-ratings and other-ratings of employees' CWB were collected. Results demonstrated that (a) self-other CWB correlations were stronger for CWB items rated by SMEs in Study 1a as more observable and less task relevant; (b) self-other CWB mean differences were smaller for CWB items rated by SMEs as more observable, public, memorable, unambiguous, and less task relevant; and (c) other-raters' overall ratings of CWB were more strongly influenced by CWB items rated by SMEs as more observable and public than were self-raters' ratings of overall CWB. In all, self-other agreement on CWB items is stronger for

items that are more observable, public, memorable, unambiguous, and less task relevant; for the most part these sorts of items are interpersonal-target CWBs (CWB-Is) rather than organizational-target CWBs (CWB-Os). These results suggest that self-other disagreement about CWB is in part due to other-raters having inadequate opportunity to observe certain CWBs (i.e., less observable and private CWBs) and having difficulty interpreting whether certain behaviors are CWBs (i.e., ambiguous, less memorable, task relevant behaviors); particularly CWB-Os. This means other-ratings of CWB-O should be viewed with skepticism and should always be supplemented with ratings from other sources, such as self-ratings.

## **DEDICATION**

First, I would like to dedicate this thesis to my parents, for all of their continued love and support. They endlessly give of themselves for the betterment of our family. “It’s not how much we give, but how much love we put into giving” (Mother Theresa). They are a true light in my life.

Second, I would also like to dedicate this project (masterpiece, really) to my siblings: Monica, Vanesa, Cris, Scarlett, Giorgina, and Rodrigo. They continue to demonstrate their love and friendship in so many ways. I know my bond with my brothers and sisters will remain strong forever, and I am happy to call them my family.

Third, I also dedicate this work to my extended family: grandparents, aunts and uncles, and cousins. God has given me a great network of support. These people are also a source of joy in daily life. Finally, this thesis is also dedicated to various friends who have stood with me in achieving this milestone.

## **ACKNOWLEDGEMENTS**

I would like to thank my committee members, Dr. Chris Berry, Dr. Kathi Miner, and Dr. Deidra Schleicher, for their guidance and feedback throughout the development of this thesis project. I would especially like to thank Dr. Chris Berry for his mentorship on this project and in my graduate school career.

I also extend my sincere gratitude to those in the program that have guided me and helped me along the path to becoming an Industrial/Organizational Psychologist, including professors, colleagues, and friends. I truly appreciate the supportive and collaborative atmosphere here, in the Texas A&M University I/O Program. Thank you to all who are a positive force in my career.

## TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
DEDICATION.....	iv
ACKNOWLEDGEMENTS .....	v
LIST OF TABLES.....	viii
1. INTRODUCTION .....	1
1.1 Background.....	1
1.2 Multi-item Measures of CWB .....	4
1.3 Self- vs. Other-Ratings of CWB .....	5
1.4 Dimension-level Classification and Predictions about What Drives Self-Other Agreement.....	8
1.5 Predictions about What Drives Overall CWB Ratings .....	18
1.5.1 Predicted Differences.....	18
1.5.2 No Predicted Differences .....	21
2. METHOD.....	23
2.1 Study 1a: Dimension Classification Ratings .....	23
2.2 Study 1b: Participants.....	24
2.3 Measures .....	27
2.3.1 Demographic Form.....	27
2.3.2 Counterproductive Work Behavior .....	27
2.4 Procedure .....	28
2.5 Analyses.....	29
3. RESULTS.....	33
3.1 Study 1a .....	33
3.2 Study 1b.....	33
3.2.1 Correlations between Dimension Ratings and Self-Other CWB Correlations .....	33
3.2.2 Correlations between Dimension Ratings and Self-Other CWB d-values ...	35
3.2.3 Correlations between Dimension Ratings and the Item-Overall CWB Correlations .....	37

4. SUMMARY AND CONCLUSIONS .....	40
4.1 Summary.....	40
4.2 Practical Implications .....	44
4.3 Limitations and Directions for Future Research.....	45
4.4 Conclusions.....	46
REFERENCES .....	48
APPENDIX A .....	51
APPENDIX B.....	53

## LIST OF TABLES

	Page
Table 1 Results of Study 1a: Dimension Scores for Each CWB Item.....	53
Table 2 Intercorrelation Matrix: Dimensions .....	54
Table 3 Intercorrelation Matrix: Study Variables.....	56
Table 4 Summary of Hypotheses, Research Questions, and Their Results .....	58
Table 5 Cohen's D for Each CWB Item.....	62
Table 6 Qualities that Characterize Each CWB Item .....	63



# 1. INTRODUCTION

## 1.1 Background

Counterproductive work behavior (CWB) is defined as intentional employee behavior that is potentially harmful and is against the legitimate interests of an organization (e.g., theft, interpersonal aggression, and poor attendance; Bennett & Robinson, 2000; Gruys & Sackett, 2003). CWB is a continuing problem for organizations in the U.S. today. For example, it is estimated that 30% to 75% of employees have engaged in some form of counterproductive behavior at work (e.g. theft or unexcused absenteeism; Harper, 1990). The presence of these behaviors can have critical negative effects on organizations. For example, costs for theft alone have been estimated to be billions of dollars to organizations annually (Camara & Schneider, 1994). In addition, employees are affected by these behaviors as well. CWBs have been linked to increased job stress and decreased job satisfaction in victims of CWB, among other things (Budd, Arvey, & Lawless, 1996).

Validly measuring any organizational behavior is difficult, but CWB measurement provides a particular challenge due to the covert nature of many CWBs. As a result, a great deal of research has been devoted to conceptualization and measurement of the CWB construct (e.g. Berry, Ones, & Sackett, 2007). First, one of the most widely-used scales of CWB over the years is the Bennett and Robinson (2000) scale. In addition, their target-based two-factor framework has been a dominant CWB framework over this period. Specifically, they proposed that CWB could be divided into two dimensions based on the target: interpersonal CWB (CWB-I; e.g., making fun of

someone or acting rudely) and organizational CWB (CWB-O; e.g., stealing items from the workplace or littering the work environment). CWB-I is deviance targeted at other individuals whereas CWB-O is deviance targeted at the organization. The items in Bennett and Robinson's (2000) CWB measure based on this framework comprised two scales: *Organizational Deviance Scale (for CWB-O)* and *Interpersonal Deviance Scale (for CWB-I)*.

However, shortly thereafter, Dalal (2005) presented meta-analytic evidence showing that CWB-I and CWB-O were highly correlated ( $r = .70$ , corrected for unreliability), which brought to question the distinction between the two types of CWB. Berry et al. (2007) examined the dimensionality of CWB further. In this meta-analysis, they concluded that evidence is suggestive of the separability of the two types of CWB based upon an examination of differential relationships with some common correlates (e.g. Big Five traits, Organizational Citizenship Behavior). Finally, another relevant work is the study by Berry, Carpenter, & Barratt (2012), which focuses on improvement of CWB measurement by examining the extent to which self-ratings (i.e. ratings made by an employee about his/her own behavior) and other-ratings (i.e. ratings by supervisors or coworkers made about an employee) of CWBs capture different information. Berry et al. (2012) presented evidence suggestive of unique information in self- and other-reports of CWB; specifically, that self- and other-reports are moderately correlated (between .38 and .44, depending on statistical corrections). Berry et al. concluded that self-ratings and other-ratings of CWB are usually viable alternatives to each other, but that they do contain sizable unique information. Additionally, it remains

unclear exactly what this unique information is and whether it is valid unique information. Therefore, there is a need to determine what are the differences and similarities in information contained in self- and other-ratings of CWB.

The present study addresses this issue in two ways. The first goal is to examine the relationship between self- and other-ratings of CWB at the item-level. Doing so will provide an understanding of the CWB items on which self-raters and other-raters agree the most. Finding the items of highest agreement between sources would suggest instances in which each rating source is basing ratings on similar information, whereas items of lowest agreement suggest instances in which rating sources are basing ratings on different information. CWB items will also be rated by subject matter experts on 10 dimensions (e.g., how observable the CWB is, the base rate of the CWB), and these item-level ratings will be used to predict and explain differences between CWB items in self-other agreement. It is important to discover the items on which various sources agree the most versus least because it will shed light on whether differences are simply due to idiosyncratic rater error (e.g., there is relatively low agreement on all CWB items) or due to some systematic factor (e.g., there is high agreement between sources on certain items and low agreement on other items, perhaps due to factors such as observability of CWBs).

The second aim of this study is to investigate the relative influence of different CWB items on overall ratings of CWB, separately for each rating source. I will examine the relationships between overall ratings of CWB and each individual item on the CWB measure. A strong relationship indicates that the item is carrying considerable weight in

the overall rating, and thus that information is being used in the overall rating. This goal contributes to measurement of CWB by discovering what information is being used by raters in overall ratings of CWB, and whether this information differs by rating source.

## **1.2 Multi-item Measures of CWB**

CWB is typically measured using multi-item measures. One example is Bennett and Robinson's (2000) measure, which was one of the first multi-item measures of CWB and has been one of the most widely used (Berry et al., 2007); it is also the multi-item CWB measure used in the present study. This particular scale was chosen for the present study for three reasons. First, it is the most prevalent CWB scale (Berry et al., 2007). Second, it is considerably shorter than other widely used scales (e.g., Spector, Fox, Penney, Bruursema, Goh, & Kessler, 2006), which was important given survey length concerns for the other-raters in the present study. Third, scores on the Bennett and Robinson (2000) measure of CWB have correlated highly with scores on other similar CWB measures (Sackett, Berry, Wiemann, & Laczo, 2006). As previously mentioned, Bennett and Robinson's (2000) instrument measures CWB according to a two-factor typology: CWB-I (measured with 7 items) and CWB-O (measured with 12 items). Raters indicate the frequency of their own engagement in the behavior (in the case of self-ratings), or the frequency with which they have observed a coworker or subordinate engage in the behavior (in the case of coworker- or supervisor-ratings): 1 (*never*), 2 (*once a year*), 3 (*twice a year*), 4 (*several times a year*), 5 (*monthly*), 6 (*weekly*), and 7 (*daily*). Thus, each item of a multi-item CWB measure represents knowledge about the frequency with which employees engage in a certain behavior. An example of a CWB-I

item is “Made fun of someone at work.” An example of a CWB-O item is “Taken property from work without permission.” Other similar multi-item measures of CWB exist and are commonly used (e.g., Spector, Fox, Penney, Bruursema, Goh, & Kessler, 2006).

The use of multi-item measures of CWB represents the perspective that there is positive manifold across the broad range of CWBs in which employees can engage. That is, engaging in one type of CWB makes it more likely that an employee will engage in other CWBs. Evidence for this positive manifold is seen in the internal consistency reliability estimates for multi-item CWB measures. For example, Berry et al.’s (2012) meta-analysis listed average alpha reliabilities ranging from .82-.89, depending on whether the multi-item measure was of CWB-I versus CWB-O and whether self-ratings versus other-ratings were used.

Perhaps most, if not all, CWB research has focused on the relationship between overall CWB (or CWB-I and CWB-O) scale scores and other variables, with little focus on item-level relationships (e.g., Penney & Spector, 2005). Therefore, little is known about what specific CWBs are driving relationships that CWB scales have with other variables; or relationships that self-report vs. other-report CWB scales have with each other.

### **1.3 Self- vs. Other-Ratings of CWB**

There are two main types of rating sources for CWB: self and others. That is, either employees rate themselves on their own counterproductivity at work, or another individual (e.g. supervisor, coworker, subordinate) rates an employee on his/her

counterproductivity at work. Self-reports are sometimes viewed with skepticism because they can be subject to issues that may compromise their validity, such as common method variance and social desirability bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Therefore, sometimes other-ratings are used instead of, or to complement, self-ratings.

Berry et al. (2012) noted that previous CWB meta-analyses (e.g. Berry et al., 2007; Dalal, 2005; Hershcovis et al., 2007) have focused mainly on self-report measures of CWB. Because of this common past limitation, their meta-analysis studied the incremental variance that other-ratings can provide beyond that of self-ratings. First, this study found that self- and other-ratings of CWB were moderately to strongly correlated with each other across samples. Specifically, the meta-analytic estimate of the correlation between self- and other-ratings was found to be .38 or .44, depending on whether alpha or inter-rater reliability is used in artifact corrections for other-ratings. Secondly, they also investigated the relationships between self- and other-ratings of CWB with common correlates (e.g. Big Five personality traits, Organizational Citizenship Behavior [OCB], organizational justice, job satisfaction). While most of the relationships self- and other-ratings had with the common correlates were similar, there were some exceptions. For example, Berry et al.'s study (2012) indicated that some interpersonal variables were more highly correlated with other-reports than self-reports (e.g. OCB-I, interactional justice, and conflict). This finding suggests that other-raters may be especially suited to rate CWBs that are interactional or interpersonal in nature (CWB-I). This would make sense because these interpersonal behaviors are more easily

observable than CWB-Os, which often tend to be carried out in private to avoid being caught (e.g., sabotage, destroying company equipment, surfing the internet instead of working). Third, the meta-analysis found that other-ratings generally accounted for little incremental variance beyond self-ratings in the common correlates. Thus, the other-ratings do not indicate much about CWB's relationships with other variables that was not already known from self-reports. Fourth, employees actually reported engaging in *more* CWBs in their self-ratings than other-raters reported those employees engaging in, suggesting that other-ratings are capturing a narrower subset of CWBs. Therefore, taken together, the results from the Berry et al. (2012) meta-analysis suggest that self-reports and other-reports contain fairly similar information, or at least there are few instances in which other-reports include valid information that is not also included in self-reports. In addition, it was proposed that, although self-reports and other-reports do contain some unique information and can be viable alternatives to each other, there is little evidence of tangible value added from collecting other-reports over self-reports, in many cases.

Although Berry et al. (2012) called into question the overall value added of other-ratings, they also highlighted some specific instances in which other-ratings may provide valid and unique information (e.g., CWB-Is). If other-ratings are more useful for certain kinds of behaviors, it is important to identify those behaviors. That is, it would be important to know if other-ratings are practical for only a subset of CWBs. The Berry et al. (2012) meta-analysis suggests that there may be more agreement between sources about CWB-I due to their public or observable nature. Specifically, they hypothesized that there would be higher agreement between self- and other-ratings for CWB-I than for

CWB-O. This hypothesis was supported and results indicated the self-other correlation for CWB-I ratings was .51 and for CWB-O ratings was .35. Therefore, I expect behaviors that are more observable to display higher agreement among sources at the item-level. In the present study, I examined the correlation between the two sources for individual CWB items. For example, for the CWB-I item “Made fun of someone at work,” the correlation between the self-rating and other-rating may be higher than the CWB-O item “Taken property from work without permission.” This example would be in line with my expectations because CWB-Is such as making fun of someone at work should be more observable to other-raters than CWB-Os such as taking property from work without permission. Stronger self-other relationships between CWB-Is than CWB-Os simply provides item-level support for the phenomenon that Berry et al. (2012) already demonstrated at the scale level. To advance understanding of *why* self- and other-raters agree/disagree for certain CWB items, a deeper and more detailed examination of the characteristics that are associated with high/low-agreement items is needed. The dimension-level analyses described in the next section address this need.

#### **1.4 Dimension-level Classification and Predictions about What Drives Self-Other Agreement**

The Berry et al. (2012) study examined the self-other agreement at the scale-level; however, it is also possible to examine self-other agreement at the item level along numerous other dimensions. The present study extends research on self-other agreement by examining 10 dimensions along which the Bennett and Robinson (2000) CWB items can be categorized. It is important to incorporate dimensions into the examination of



self-other agreement because each dimension represents a quality that characterizes CWBs (e.g. how memorable the CWB is or how observable the CWB is). Each CWB item is characterized by how high or low it falls on each of the 10 dimensions.

Therefore, each CWB item has 10 qualities or traits that describe that item. These qualities can shed light on the particular nature of self-other agreement for each item.

Each dimension ultimately represents a potential reason for disparity of knowledge about CWB between sources. For example, the observability dimension (presented later in this section) ultimately illustrates a potential reason for one source (e.g., self-rater) knowing something about CWB that the other source (e.g., other-rater) would not know. Namely, the self would have knowledge of engagement in CWB that the other-rater would not have due to *observability* (i.e., whether or not a behavior is visible). This is true for each dimension. Namely, each dimension represents a potential reason for one source having knowledge about CWB that the other source does not have access to. Therefore, if the dimension illustrates a discrepancy of knowledge about the CWB between sources (e.g., the self-rater knows something about their own CWB that the other-rater does not know), self-other agreement will potentially be affected. Because classification of CWB items along dimensions is a new research area, there is relatively little research on it. However, there are some notable studies that have contributed to the dimensions I present here. As previously mentioned, the major distinction among CWB items has been CWB-I versus CWB-O (Bennett and Robinson, 2000). However, the present study seeks to extend this research line, so the focus is on new dimensions.

First, Observability is a central quality to CWB behaviors. This dimension refers to the extent to which each particular CWB is likely to be observed. The Realistic Accuracy Model (RAM; Funder, 1995) can shed light on this dimension as well as a few others. The RAM Model proposes that accurate judgment of personality occurs through the following path: relevance of behavioral cues, availability for observation, detection, and utilization of relevant behavioral cues. Although this model is generally used in the area of personality ratings, it can be applied to CWB ratings as well. To illustrate the model with an example: first, there must be a relevant behavior (e.g., an employee is stealing a printer), second, the behavior must be available for observation (e.g., a coworker or supervisor must be able to see it), third, the behavior must be detected (e.g., in addition to being able to see it, a coworker or supervisor must actually see it), and finally, the behavior must be interpreted properly (e.g., the coworker or supervisor must conclude that the employee committed theft and not something else, such as borrowing the printer).

According to the second step of the RAM pathway, “availability for observation” (Funder, 1995) implies that cues must be literally available for observation. CWB’s that are more observable may exhibit higher agreement between sources than CWB’s that are not as observable. This is because other-raters must be able to observe the behavior in order to recall it at the time of rating. Self-raters are more likely to recall a particular CWB in which they engaged regardless of observability because they experienced the CWB first-hand. This illuminates a particular discrepancy of knowledge between sources. Namely, when unobservable CWBs are engaged in, it is likely the self-rater has

more knowledge than the other-rater. I predict that higher agreement will occur for observable behaviors because it is more likely that both sources will have access to knowledge about observable behaviors as opposed to less observable ones. Higher agreement is expected to manifest itself in terms of smaller mean differences between self- and other-ratings as well as higher self-other item-level correlations.

*Hypothesis 1a: Self-other correlations will be stronger for more observable CWBs.*

*Hypothesis 1b: Mean differences will be smaller for more observable CWBs.*

Second, the public vs. private distinction is also relevant to CWB dimensionality. Oh et al. (2014) found evidence through a survey of SME's suggesting a difference in setting for CWB-I vs. CWB-O. Specifically, their findings indicate that CWB-I's are more likely to occur in public settings while CWB-O's are more likely to occur in private settings. This distinction is also applicable to the goals of this study. It is possible that public behaviors are more likely to display higher agreement between sources than private behaviors. This is based on a similar reasoning to the observability dimension, but they are not interchangeable. The observability dimension focuses on the likelihood that a supervisor/coworker would simply observe a particular CWB in action. The public/private dimension focuses on the CWB itself and whether or not its engagement tends to occur in a public or private place. Therefore, it is possible that a CWB is observable but tends to occur in a private setting (e.g., littering your work environment). I predict that higher agreement between sources will be seen for behaviors that are public. This is because both sources are more likely to witness public behaviors (i.e. public CWB) than private behaviors (i.e., private CWB).

*Hypothesis 2a: Self-other correlations will be stronger for more public CWBs.*

*Hypothesis 2b: Mean differences will be smaller for more public CWBs.*

Third, I expect to find that Memorability will influence self-other agreement of CWB. According to the RAM Theory (Funder, 1995), the more accessible the cue is, the more likely it will be recalled and utilized correctly in judgment of personality. I believe the same is true for CWB. Namely, Memorability would positively influence recall and utilization of cues relevant to judgment of a person as it relates to engagement in CWB. During the rating process, it seems likely that other-raters would recollect memorable items at a higher capacity than unmemorable items. Self-raters may recollect the CWB regardless of memorability because they actually engaged in it; thus, they generally have access to a richer experience. Other-raters must rely directly on memory of what was observed when rating CWBs. I expect higher agreement between sources for memorable items than unmemorable items.

*Hypothesis 3a: Self-other correlations will be stronger for more memorable CWBs.*

*Hypothesis 3b: Mean differences will be smaller for more memorable CWBs.*

Next, the first step of the RAM pathway (i.e., relevance of behavioral cues) implies that cues must be relevant to a particular behavior for accurate personality judgment (i.e., the cues should not be ambiguous). The same is true for CWB judgment. Namely, cues that indicate voluntary CWBs should be clear enough that they are unlikely to be confused with another behavior/trait. For example, it may be the case that intentionally working slowly can be easily confused with a lack of skill, whereas the cues for verbal abuse may be less ambiguous. I predict that cues that are less likely to be

indicative of multiple behaviors (i.e., cues that are clear/unambiguous) will have higher agreement than ambiguous cues. If a cue can be misinterpreted as originating from things other than CWBs, then the other-rater may not correctly label a particular behavior as CWB. However, the self-rater would know the correct interpretation of the cue. This would be an example of the disagreement that could occur between sources due to this dimension. Therefore, CWBs that have clear cues will have higher agreement among sources.

*Hypothesis 4a: Self-other correlations will be stronger for more unambiguous (clear) CWBs.*

*Hypothesis 4b: Mean differences will be smaller for more unambiguous (clear) CWBs.*

Fifth, Social Desirability measures how socially desirable it is to engage in each CWB. In many situations, people engage in socially desirable behavior, which contributes to making a positive impression on others (Tetlock & Manstead, 1985). It is possible that social desirability influences the agreement between self and others for CWB. One possibility is that the less socially desirable a CWB, the more likely an employee is to hide it. In the same way, employees may not be as careful to hide behaviors that are not as socially undesirable. I predict that higher agreement will be found for behaviors that are more socially desirable than those that are more socially undesirable.

*Hypothesis 5a: Self-other correlations will be stronger for more socially desirable CWBs.*

*Hypothesis 5b: Mean differences will be smaller for more socially desirable CWBs.*

For the remaining dimensions (Base Rates, Severity, Stigma, Organizational Sanctions, and Task Relevance), there are competing possibilities for how these attributes of CWBs will affect self-other agreement. Therefore, the influence of these dimensions on self-other agreement about CWB will be presented as research questions.

Robinson and Bennett (1995) first proposed the sixth dimension of the present study, namely, whether the CWB is minor (e.g. coming in late) or more serious in nature (e.g. using drugs; I refer to this dimension as Severity). It is not difficult to imagine that severity of CWB may influence the extent to which others (e.g. supervisors and coworkers) and the self agree at the item-level. For example, it is possible that the more serious CWBs are hidden more effectively. This may be because employees engaging in severe CWBs have a strong motive to hide their CWBs; namely, they want to avoid the serious punishment that is associated with them. Thus, it is possible that there will be less self-other agreement about more severe CWBs because of the lengths to which employees would go to hide such CWBs. However, severe behaviors also tend to be more noticeable and difficult to hide, because their very nature likely causes a serious outcome (e.g., consuming drugs or alcohol on the job may have noticeable, physical effects). Therefore, it is also possible that severe behaviors would lead to higher agreement between sources. Because there are competing explanations for this dimension's effect on self-other agreement, this prediction is exploratory.

*Research Question 1a: How does Severity influence self-other correlations of CWBs?*

*Research Question 1b: How does Severity influence self-other mean differences of CWBs?*

Seventh, Gruys and Sackett (2003) previously examined a dimension involving task vs. non-task behaviors. They found that there is a distinction between CWBs that are task-related (e.g., intentionally working slowly on job tasks) and task-unrelated (e.g., verbal abuse; this dimension is named Task Relevance). Bowling and Gruys (2010) pointed out that this dimension has been previously under-explored or overlooked. It is possible that other-ratings would be more in agreement with self-ratings for either the task or non-task behaviors due to competing explanations. One possibility is that higher agreement may occur for task behaviors because other-raters may have a motive to pay attention to employees' task-related CWBs. For example, supervisors may have a need to remember task-related CWBs when considering employees for promotions or raises. Similarly, coworkers may rely on each other to complete various job tasks, making task CWBs especially salient in the workplace. However, it is also possible that higher agreement occurs for non-task behaviors. Similar to the ambiguity dimension, task-related CWBs may be more ambiguous to other-raters (e.g., is it CWB or just incompetence?) Under this possibility, the non-task behaviors would be less ambiguous, thus facilitating agreement. Therefore, due to competing explanations, examining the effect of this dimension on self-other agreement is also exploratory.

*Research Question 2a: How does Task Relevance influence self-other correlations of CWBs?*

*Research Question 2b: How does Task Relevance influence self-other mean differences of CWBs?*

Eighth, according to Bowling and Gruys (2010), little attention has been given to differing base rates among CWBs as well. This dimension is included in the present list because I believe that the sheer number of times a particular CWB is committed on average may influence self-other agreement (this dimension is referred to as Base Rates). There are more opportunities to observe a behavior that has a high base rate, regardless of whether the setting is generally public or private. This possibility would increase agreement across sources. However, a higher base rate may indicate that a behavior is either 1) often overlooked, 2) minor in nature, or 3) more well-hidden due to more employee engagement. This possibility would predict lower agreement across sources. Therefore, it is not clear if higher base rates are linked to more agreement between sources, and examination of dimension is exploratory.

*Research Question 3a: How does Base Rates influence self-other correlations of CWBs?*

*Research Question 3b: How does Base Rates influence self-other mean differences of CWBs?*

Ninth, another dimension that was included is referred to as Organizational Sanctions. This dimension examines how strong the punishment is likely to be for each particular CWB if the employee is caught in the act. It is possible that the level of punishment influences the extent to which there is agreement between supervisors/coworkers and employees. One possibility is that employees are more likely



to conceal a behavior if the punishment is more serious, which would lead to lower agreement among sources. However, similar to the severity dimension, the competing explanation is that CWBs with more organizational sanctions may be more difficult to hide and be more noticeable. This would positively impact the agreement across sources. Therefore, this dimension is exploratory.

*Research Question 4a: How does Organizational Sanctions influence self-other correlations of CWBs?*

*Research Question 4b: How does Organizational Sanctions influence self-other mean differences of CWBs?*

Tenth, the Stigma dimension measures the extent to which the CWB is characterized by social stigma (i.e., a negative, social perception attached to an object). This dimension is examined in the present study because I believe that the stigma of a behavior may affect self-other agreement. Because a behavior with a stigma is characterized by a negative perception, this perception is likely to be attached to the person who committed the CWB. Most, if not all, CWBs have some stigma attached to them. However, there are likely differences between CWBs in the amount of stigma. For example, using illegal drugs on the job entails more stigma than intentionally putting little effort into the work or gossiping about coworkers. Similar to other dimensions, there are competing possibilities regarding effects of stigma of CWBs on self-other agreement. First, it is possible for the presence of stigma to be positively related to agreement between sources because the stigmatized behavior is more likely to be recalled than unstigmatized behaviors by raters during the rating process. However, I

also expect that doers of CWBs are aware of stigmatization of behaviors, which may prompt doers to hide behaviors more effectively. Because of the two competing possibilities, this dimension is exploratory as well.

*Research Question 5a: How does Stigma influence self-other correlations of CWBs?*

*Research Question 5b: How does Stigma influence self-other mean differences of CWBs?*

### **1.5 Predictions about What Drives Overall CWB Ratings**

The previous set of hypotheses and research questions focused on the prediction of what drives self-other agreement among sources for CWB. Another goal of this study is to examine differences in what drives the overall CWB rating for each source. Therefore, questions I am interested in answering include: “What qualities (i.e., dimensions) are most predictive of similarities between sources in overall CWB ratings?” and “What qualities are most predictive of differences between sources in overall CWB ratings?” The next set of hypotheses address this issue.

**1.5.1 Predicted Differences.** I expect certain dimensions of CWB items to predict differences in how sources arrive at an overall CWB rating. The first dimension that is expected to predict differences in how each source arrives at its respective overall CWB rating is Social Desirability. Self-raters will perhaps be less likely to admit to socially undesirable CWBs because they are engaging in impression management. Other-raters do not have such social desirability concerns about reporting on the CWBs of others. Therefore self-raters have a motive to underreport more socially undesirable

CWBs that other-raters do not have. This means that more socially undesirable behaviors will influence the overall other CWB rating more than the overall self-rating.

*Hypothesis 6: More socially undesirable items are expected to influence the overall other-rating more than the overall self-rating.*

The next dimension that is expected to predict differences in overall CWB is Ambiguity. Ambiguity refers to the level of ambiguity of cues for CWBs to any observer of the CWB (i.e., an other-rater). It is likely that the self-rater is aware of whether or not a behavior is a CWB. However, a typical other-rater does not have access to knowledge clarifying whether or not a particular ambiguous behavior is clearly a CWB. That is, there is a disparity of information between sources. Therefore, ambiguous CWBs are expected to influence self-raters' overall CWB rating more than other-raters' overall CWB rating. This is because self-raters tend to know if these ambiguous CWBs are in fact CWBs. However, other-raters would generally not use ambiguous CWB information in the overall rating if they believe the CWB was too ambiguous or likely caused by something else. Therefore, other-raters are likely to disproportionately rely on the information provided by the clear instances of employee CWB, rather than the unsure or ambiguous ones.

*Hypothesis 7: More unambiguous items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB.*

The next dimension that is expected to predict differences in how each source arrives at the overall CWB rating is Task Relevance. Similar to the ambiguity dimension, task-related CWBs may be more ambiguous to other-raters. In general, there

are at least two explanations for a task behavior being performed sub-optimally. One reason is that the employee could be intentionally performing at a lower level (indicating CWB). Second, the employee may be unintentionally performing at a lower level (indicating potential lack of training, ability, skill or knowledge). For example, other-raters may not know if an employee is intentionally working slowly (task-related CWB) or is incompetent with a particular task, thus making task behaviors more ambiguous. However, since task-unrelated behaviors do not necessarily require particular training or skills (e.g., being courteous or respecting others), task-unrelated counterproductivity is less ambiguous. I propose that other-raters would disproportionately rely on information that denotes CWB clearly as opposed to information that is ambiguous. Therefore, other-raters are likely to rely more on non-task CWBs, which are not as ambiguous as task behaviors.

*Hypothesis 8: Task-unrelated items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB.*

The next set of dimensions expected to predict differences in overall CWB is the observability and public/private dimensions. Observable CWBs are likely to influence the other-rating of overall CWB more than the self-rating of overall CWB. This is because other-raters must rely on behaviors they observe employees engaging in. However, self-raters have relevant knowledge of observable CWBs and unobservable CWBs because they are aware of what behaviors they engaged in. There is a discrepancy in knowledge between sources. Therefore, other-raters are limited to use only what they know, namely, observable CWBs. In the same way, public behaviors are

expected to drive the other-rating of overall CWB more than the self-rating of overall CWB. Again, other-raters are more likely to be aware of employees engaging in public behaviors than private behaviors.

*Hypothesis 9: More observable items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB rating.*

*Hypothesis 10: More public items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB.*

Finally, Memorability is also expected to predict differences in how each source arrives at their respective overall CWB rating. Memorable items should influence the overall other-rating more than the overall self-rating because other-raters have to rely more on what they can remember from observation. However, self-raters know what they did regardless if others saw it or not. That is, self-raters have experience and memory, while other-raters have to rely on memory alone. Therefore, there is a natural discrepancy of knowledge that results from self-raters actually experiencing their CWB as opposed to other-raters simply observing and subsequently remembering the CWB.

*Hypothesis 11: More memorable items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB.*

**1.5.2 No Predicted Differences.** There are a number of dimensions (Severity, Base Rates, Organizational Sanctions, and Stigma), for which there is not a clear rationale for expecting them to differentially affect the relationship between CWB items and overall CWB for self- versus other-ratings of CWB. For example, for CWBs that have a high base rate, I believe other-raters' and self-raters' overall CWB rating will be

influenced in a similar manner. The same is expected for CWBs that have a low base rate. The amount of times a CWB tends to occur in general is a shared perception by all sources. There is no reason the other-rater or the self-rater would be disproportionately influenced by the base rate when assessing overall CWB. Therefore, no hypotheses are made about the effects of these dimensions on the relationship between individual CWB items and overall CWB. However, the effects of these dimensions will be investigated in the present study in a purely exploratory fashion.

## 2. METHOD

### 2.1 Study 1a: Dimension Classification Ratings

The purpose of Study 1a is to obtain from subject matter experts (SMEs) ratings of each CWB item on the 10 dimensions; these CWB dimension ratings for each item will then be used in Study 1b to predict self-other agreement for each CWB item. A Qualtrics survey was used to collect dimension classifications. SMEs classified each CWB item along the 10 dimensions presented. SME's were 9 graduate students in the Industrial/Organizational Psychology Ph.D. program at a large Southwestern university. The survey included the descriptions of the 10 dimensions as well as a rating form for SME's to classify each CWB item along each dimension. For example, the instructions for the Observability dimension read "How observable is this CWB to supervisors and coworkers?" Ratings were given on a 7-point Likert scale for each dimension (e.g., 1 = not at all observable, and 7 = very observable). This includes Public/Private (e.g., 1 = always in a private setting, and 7 = always in a public setting), Organizational Sanctions (1 = no punishment, and 7 = very strong), Base Rates (1 = never, 2 = once a year, 3 = twice a year, 4 = several times a year, 5 = monthly, 6 = weekly, 7 = daily), Stigma (1 = not at all stigmatized, 7 = very stigmatized), Memorability (1 = very unmemorable, 7 = very memorable), Severity (1 = not at all severe, 7 = very severe), Social Desirability (1 = very socially undesirable, 7 = very socially desirable), Ambiguity (1 = very ambiguous, 7 = very clear), and Task Relevance (1 = completely unrelated, 7 = completely related).

Interrater reliabilities ranged from .84 to .97 for the 190 dimension classifications (i.e., 19 items by 10 dimensions). Despite these high interrater reliabilities (which are influenced, in part, by the large number of raters: 9 raters), there were some dimension ratings with relatively large standard deviations (operationalized as a SD of 1.8 or higher). There were two types of items with such large SDs: those characterized by one outlier rating (at least 2 SDs above or below the mean), and those characterized by split ratings, with multiple raters rating a CWB items as high on a dimension at the same time that multiple raters rated the same item as low on a dimension. When the large SD was due to one outlier rater, the outlier was deleted for that item-dimension rating. However, in instances wherein the large SD was due to a more even split, the divergent ratings were kept as they likely represented genuine differences in opinion among the SMEs.

## **2.2 Study 1b: Participants**

The purpose of Study 1b was to collect self- and other-ratings of employees' CWB and to use the CWB item dimension ratings from Study 1a to predict agreement between self- and other-ratings of CWB. There were two types of participants in Study 1b: employees (self-raters) and supervisors/coworkers (other-raters). In line with past research, supervisors and coworkers were grouped together as "other-raters" based on evidence that supervisor- versus coworker-ratings do not moderate the self-other CWB relationship or the relationships that CWB has with other variables (e.g., Berry et al., 2012). 85 pairs of participants (i.e., an employee paired with either a supervisor or coworker) were recruited, which is the minimum sample required by the power analysis to detect an effect. Specifically, according to Cohen's (1992) power analysis tables, 85



participants are required to have power equal to .80 to detect a medium effect size for the bivariate correlations between self- and other-report CWB items (i.e. the analysis that requires the most participants; testing mean differences between self- and other-reports only requires 64 participants, according to Cohen, 1992). I expect a “medium effect size”, as the meta-analytic correlation between self- and other-reports of CWB at the scale-level is approximately .40, a medium-to-large effect size, according to Cohen (1992). Employee participants were recruited from an online task-based website, Amazon’s Mechanical Turk (M Turk). M Turk is a website where participants log in and choose to take surveys or do tasks from their own computer. Each employee was asked to provide their own first name and last initial (e.g., Juan B.) and the email address of their supervisor or coworker, who will then be recruited through email individually. I asked the employee to create a secret code, which was then used to link each employee to their corresponding supervisor/coworker. In the recruitment email, supervisors/coworkers were given the first name and last initial of the employee as well as the secret code the employee created. When supervisors/coworkers filled out their surveys, they entered that code. Secret codes were linked in the data file to identify each employee with their corresponding supervisor/coworker. The data are only interpretable if employees are correctly linked to their supervisors/coworkers. The secret codes are necessary to avoid handling personally identifiable information (i.e., full names).

For inclusion purposes, employee participants had to be users of Amazon’s Mechanical Turk. In addition, they had to have a job in which they work at least 20 hours per week as well as be able to provide contact information for their supervisor

and/or coworker(s). After successful completion of the given tasks, the self-raters then received payment of 50 cents; supervisor/coworker raters were entered into a drawing for a \$50 Amazon.com gift card.

My study required a second wave of data collection via a different method: Study Response<sup>1</sup>. Study Response (<http://www.studyresponse.net/index.htm>) is an online data collection service that has proven to be a valuable tool in past social science research (e.g., Wallace, 2004). For a fee to the researchers, Study Response connects the researchers with participants who have signed up with Study Response specifically in order to participate in online research. For inclusion purposes, participants had to be contacted by Study Response. Participants recruited from this method still engaged in the same basic survey procedures as participants recruited from M Turk. However, study response participants did not have to create a secret code because Study Response recruited them directly. In addition, because of the same reason, there was no need for the recruitment email for supervisors/coworker for this group of participants.

---

<sup>1</sup> After collecting approximately 60 self-supervisor/coworker pairs, M Turk asked that data collection stop because they felt this study violated their terms of agreement by collecting individually-identifiable data about participants. Thus, 25 additional pairs of data were needed and the present study used Study Response to collect these 25 pairs. The two subsamples did not differ in age or hours worked per week. However, they did differ on age and CWB. The Study Response sample was older than the M Turk sample, and the Study Response sample engaged in less CWB. It is no surprise that older employees tend to engage in less CWBs. This may even be a good thing if we look at it from the perspective that if anything the Study Response sample would attenuate correlations due to lower reporting of CWB.

## 2.3 Measures

**2.3.1 Demographic Form.** Participants provided basic information such as age, gender, and race/ethnicity, as well as various characteristics about their job (e.g., industry, length of time employed).

**2.3.2 Counterproductive Work Behavior.** CWB was assessed using Bennett and Robinson's (2000) 19-item measure of Workplace Deviance. Specifically, it includes 7 items measuring Interpersonal Deviance and 12 items measuring Organizational Deviance. An overall item of CWB was created for this study and included as well, "Overall, how often do you engage in counterproductive work behaviors?" This item was created in order to examine what particular items are driving an overall rating of CWB. Participants' responses indicate frequency of engaging in each CWB on a 7-point scale. The scale consists of the following: 1 (*never*), 2 (*once a year*), 3 (*twice a year*), 4 (*several times a year*), 5 (*monthly*), 6 (*weekly*), and 7 (*daily*). Reliability using Cronbach's alpha for the self-report measure was .93 and for the other-report measure was .98.

Both self-raters and other-raters filled out the CWB measure. The original CWB measure was worded from the self-rater perspective. Because of this, different wording was used for the other-raters so that they could fill it out from their perspective. For example, the instructions of the self-rater measure read "How often have you engaged in the following behaviors?" while the instructions of the other-rater measure read "How often has your employee or coworker engaged in the following behaviors?"

## 2.4 Procedure

Participants were recruited through M Turk. When participants log in to their M Turk account, they are free to choose to participate in any survey or task they want. When my survey became available, one of their options was my workplace survey for this study. If they chose to take my survey, they clicked on the survey link on M Turk, which took them to the survey, which is on Qualtrics.com. The participants affirmed the consent form online and then read the instructions for the survey. Before filling out any of the study measures, participants were asked to provide contact information (e.g., an email address) for a supervisor and/or coworker, their own first name and last initial (this was used when contacting supervisor/coworker raters so that they knew who they were supposed to be rating), and were asked to create a secret code that I later used to anonymously link participant responses to their respective supervisor/coworker ratings. Then participants filled out the study measures of interest. After filling out the survey, Qualtrics provided participants with a code number that they could enter in M Turk to receive their payment.

After participants completed the survey, I emailed the supervisor/coworker of the participant listed, asking the supervisor/coworker to fill out a survey about the participant. Only the first name and last initial of the participant, along with the participant's secret code, was provided to supervisor/coworker raters. The supervisor/coworker raters was asked to rate the participant on the same CWB scale that participants completed. The supervisor/coworker raters were also be asked to enter the participant's secret code, which was used to anonymously link supervisor/coworker

ratings to their respective participant ratings. Supervisors/coworkers were given the opportunity to enter a drawing for a \$50 Amazon.com gift card as incentive for filling out the questionnaire. Any contact information provided for the prize drawing was kept separate from the supervisor/coworker ratings data.

The same procedures apply to the second wave of data collection. However, there was no need for the secret code protocol to link employees and supervisors/coworkers because Study Response took care of this. In addition, there was no need for the payment code because Study Response paid participants directly.

## **2.5 Analyses**

For the first set of hypotheses, I ran two analyses. First, in order to examine the disagreement/convergence between sources at the item-level, I calculated bivariate correlations between self- and other-ratings for each item (collected in Study 1b) across all dimensions (dimension ratings taken from Study 1a). Five dimensions are examined as hypotheses (i.e., Observability, Public/Private, Memorability, Ambiguity, and Social Desirability), while the other five dimensions are examined as research questions (i.e., Severity, Task Relevance, Base Rates, Organizational Sanctions, and Stigma). In order to examine if the dimensions were predictive of self-other CWB agreement, I calculated bivariate correlations between each dimension rating from Study 1a (predictor) and the self-other CWB bivariate correlations from Study 1b. If a given dimension has an effect on self-other agreement, I would expect there to be a correlation between that dimension and the self-other correlations. For these correlations between dimensions and self-other correlations, the sample size was limited to 19, as that is the number of CWB items.

Given this practical constraint and its effects on statistical power, an alpha level of .10 will be used for these correlations, rather than the typically (and arbitrarily) used alpha level of .05. In this case, the critical value of these correlations for statistical significance is .389.

Specifically, these bivariate correlations will be used to examine the following set of hypotheses and research questions: (1a) Self-other correlations will be stronger for more observable CWBs, (2a) Self-other correlations will be stronger for more public CWBs, (3a) Self-other correlations will be stronger for more memorable CWBs, (4a) Self-other correlations will be stronger for more unambiguous (clear) CWBs, (5a) Self-other correlations will be stronger for more socially desirable CWBs, (RQ1a) How does Severity influence self-other correlations of CWBs, (RQ1b) How does Severity influence self-other mean differences of CWBs, (RQ2a) How does Task Relevance influence self-other correlations of CWBs, (RQ2b) How does Task Relevance influence self-other mean differences of CWBs, (RQ3a) How does Base Rates influence self-other correlations of CWBs, (RQ3b) How does Base Rates influence self-other mean differences of CWBs, (RQ4a) How does Organizational Sanctions influence self-other correlations of CWBs, (RQ4b) How does Organizational Sanctions influence self-other mean differences of CWBs, and (RQ5a) How does Stigma influence self-other

correlations of CWBs, (RQ5b) How does Stigma influence self-other mean differences of CWBs?<sup>2</sup>

Regarding the set of hypotheses aimed at self-other mean differences in CWB ratings, I first calculated *t*-tests between self- and other-ratings for each CWB item to determine on which items there were significant differences. Then, standardized mean differences (*d*-values) between self-ratings and other-ratings were calculated for each CWB item. These item-level CWB *d*-values then acted as criteria in analyses wherein the *d*-values were correlated with each of the 10 CWB dimensions from Study 1a in order to test whether the dimension ratings influenced the size of the self-other mean difference in CWB ratings. These latter analyses will be used to examine the following hypotheses: (1b) Mean differences will be smaller for more observable CWBs, (2b) Mean differences will be smaller for more public CWBs, (3b) Mean differences will be smaller for more memorable CWBs, (4b) Mean differences will be smaller for more unambiguous (clear) CWBs, and (5b) Mean differences will be smaller for more socially desirable CWBs. The same will be done for the exploratory dimensions (i.e., Research Questions 1a-5b).

For the set of hypotheses regarding what drives self- versus other-ratings of overall CWB, I first calculated bivariate correlations between each CWB item and the

---

<sup>2</sup> It would be ideal to test these hypotheses and research questions using multiple regression instead of bivariate correlations. However, the “sample size” for such a multiple regression analysis would be 19, a practical constraint imposed by the number of CWB items. A sample size of 19 is not sufficient for a multiple regression analysis containing 10 predictors (the 10 dimensions), so bivariate correlations were the only option. This also applies to the other analyses described below (e.g., the analyses predicting self-other mean differences).

overall rating for each source to arrive at the CWB item-overall CWB correlation for each source. Therefore, this resulted in 19 correlations for self-raters (self CWB item-overall correlations) and 19 correlations for other-raters (other CWB item-overall correlations). Then, to test whether or not the dimensions predict differences in how each source arrives at the overall CWB rating, dimension ratings were correlated with each source's CWB item-overall CWB correlations. This set of analyses was used to examine the following hypotheses: (6) More socially undesirable items are expected to influence the overall other rating more than the overall self-rating, (7) More unambiguous items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB, (8) More task-unrelated items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB, (9) More observable items are expected to influence the other-rating of overall CWB rating more than the self-rating of overall CWB rating, (10) More public items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB, and (11) More memorable items are expected to influence the other-rating of overall CWB rating more than the self-rating of overall CWB.



### 3. RESULTS

#### 3.1 Study 1a

Table 1 presents the results from Study 1a. Each row in the table refers to a particular CWB item, while each column refers to one of the ten dimensions examined in the present study. Each item has a dimension score for each dimension, which is represented by the mean (i.e., of SME ratings). Therefore, each of Bennett and Robinson's (2000) 19 items has 10 dimension scores (one for each dimension examined). The standard deviation is shown below the mean for each item-dimension combination.

Table 2 shows an intercorrelation matrix of dimension ratings for all dimensions. This shows how correlated the dimension ratings were to each other. For example, Observability and Public/Private dimensions are significantly correlated ( $r = .84$ ). This makes sense because observable behaviors tend to be public as well. However, dimensions which have no theoretical reason to be related should show a relatively minor or no correlation. For example, Observability and Memorability are not significantly correlated ( $r = .27, p > .05$ ), because behaviors that are observable are not necessarily memorable.

#### 3.2 Study 1b

##### *3.2.1 Correlations between Dimension Ratings and Self-Other CWB*

**Correlations.** Table 3 lists the intercorrelations between all study variables. Of particular interest are the item-level correlations between the self- and other-ratings of CWB listed on the diagonal. Self-other correlations range from .153 to .589 across the

20 CWB items, supporting the idea that sources agree more about some CWB items than others.

For Hypothesis 1a, the observability dimension score and self-other agreement were significantly correlated ( $r = .51, p < .05$ ), as can be seen in Table 4. This indicates that the more observable the CWB, the higher the self-other agreement. Thus, hypothesis 1a was supported (see also Table 4 for a summary of the results and interpretations for each hypothesis test and research question). For this correlation as well as every other one, tests of significance were two-tailed.

The correlation between the public/private dimension and self-other agreement was not significant ( $r = .31, p > .10$ ). Hypothesis 2a was not supported. The correlation between the memorability dimension and self-other agreement was not significant ( $r = .29, p > .10$ ). Hypothesis 3a was not supported. The correlation between the ambiguity dimension and self-other agreement was not significant ( $r = .36, p > .10$ ). Hypothesis 4a was not supported. The correlation between the social desirability dimension and self-other agreement was not significant ( $r = -.05, p > .10$ ). Hypothesis 5a was not supported.

In response to RQ2a, the correlation between Task Relevance and self-other agreement was significant ( $r = -.50, p < .05$ ). This indicates that the more task-unrelated the CWB, the more self-other agreement. Although formulated as a research question from the start, this finding is significant and sizeable.

For the remaining research questions, correlations between dimensions and self-other agreement were not significant. The correlation between severity and self-other

agreement was not significant ( $r = .14, p > .10$ ; RQ1a). The correlation between base rates and self-other agreement was not significant ( $r = .09, p > .10$ ; RQ3a). The correlation between organizational sanctions and self-other agreement was not significant ( $r = .09, p > .10$ ; RQ4a). Finally, the correlation between stigma and self-other agreement was not significant ( $r = .17, p > .10$ ; RQ5a).

### ***3.2.2 Correlations between Dimension Ratings and Self-Other CWB $d$ -values.***

20 independent sample  $t$ -tests were analyzed comparing self-report CWB and other-report CWB for each of the 20 individual items (please refer to Table 5 for results). Independent sample  $t$ -tests indicated significant differences between self and other ratings (self-ratings higher for each) for item 9 ( $t(188) = 2.24, p < .05, d = .33$ ), item 11 ( $t(188) = 3.31, p < .01, d = .48$ ), and item 18 ( $t(188) = 2.23, p < .05, d = .32$ ). These items refer to “Spent too much time fantasizing or daydreaming instead of working,” “Taken an additional or longer break than is accepted in your workplace,” and “Put little effort into your work” respectively. These  $t$ -tests indicate that the largest amount of differences between sources were for these three items.  $d$ -values were also calculated for each item to supplement the  $t$ -tests (see Table 5) and to use as criteria in the next set of analyses.

Next, the dimension ratings from Study1a were correlated with the item-level  $d$ -values from Study 1b to investigate which dimensions are predictive of self-other mean differences in CWB ratings. For Hypothesis 1b, the observability dimension score and  $d$ -values were significantly correlated ( $r = -.50, p < .05$ ). This indicates that the more

observable the CWB, the less mean differences between self and others (indicating higher self-other agreement). Thus, hypothesis 1b was supported.

The correlation between the public/private dimension and  $d$ -values was significant ( $r = -.51, p < .05$ ). This correlation means that the more public a CWB, the smaller the mean differences between sources. Therefore, hypothesis 2b was supported.

The correlation between the memorability dimension and  $d$ -values was significant ( $r = -.49, p < .05$ ). This correlation means that the more memorable a CWB, the smaller the mean differences between sources. Hypothesis 3b was supported.

The correlation between the ambiguity dimension and  $d$ -values was significant ( $r = -.42, p < .10$ ). This correlation means that the more clearly a behavior denotes CWB (i.e., is less ambiguous), the smaller the mean differences. Hypothesis 4b is supported.

The correlation between the social desirability dimension and  $d$ -values was not significant ( $r = .32, p > .10$ ). Hypothesis 5b is not supported.

In response to RQ1b, the correlation between severity and  $d$ -values was significant ( $r = -.42, p < .10$ ). This means that the more severe the CWB, the smaller the mean differences between sources.

In response to RQ2b, the correlation between Task Relevance and  $d$ -values was significant ( $r = .48, p < .05$ ). This indicates that the more task-unrelated the CWB, the less mean differences. Like the correlation with self-other agreement, this correlation is also significant.

The correlation between base rates and  $d$ -values was sizeable but not significant ( $r = .33, p > .10$ ). RQ3b was not supported. The correlation between organizational

sanctions and self-other agreement was not significant ( $r = -.05, p > .10$ ). RQ4b was not supported.

Finally, in response to RQ5b, the correlation between stigma and  $d$ -values was significant ( $r = -.50, p < .05$ ). This means that the more stigma attached to the behavior, the smaller the mean differences, indicating higher self-other agreement.

### ***3.2.3 Correlations between Dimension Ratings and the Item-Overall CWB***

***Correlations.*** For hypothesis 6, the correlation between social desirability and other-rater item-overall correlations was not significant ( $r = .28, p > .10$ ). However, the correlation between this dimension and self-rater item-overall correlation is significant ( $r = .46, p < .10$ ). I predicted that social undesirability would influence the other overall rating more than the self-rater overall rating, so hypothesis 6 was not supported.

For hypothesis 7, the correlation between ambiguity and other item-overall correlations was not significant ( $r = -.16, p > .10$ ). The correlation between ambiguity and self-rater item-overall correlations was also not significant ( $r = -.37, p > .10$ ). Since I predicted that less ambiguous items would influence the overall other-rating more than the overall self-rating, hypothesis 7 is not supported.

Hypothesis 8 indicated that more task-unrelated items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB. The correlation between Task Relevance and other item-overall correlations was not significant ( $r = .24, p > .10$ ). The correlation between Task Relevance and self-rater item-overall correlations was not significant either ( $r = -.05, p > .10$ ). Hypothesis 8 was not supported.

Hypothesis 9 indicated that more observable items are expected to influence the other-rating of overall CWB rating more than the self-rating of overall CWB rating. The correlation between Observability and other item-overall correlations was significant, ( $r = .50, p < .05$ ). This means that the more observable the CWB item, the more it was used in arriving at the other overall CWB rating. The correlation between Observability and self-rater item-overall correlations was very small and not significant ( $r = .03, p > .10$ ). This indicates that Observability does not influence how much the item is used in the overall rating for self-raters. Taking both correlations together, they indicate that more observable items influence the other overall rating more than the self-rater overall rating. Hypothesis 9 was supported.

Hypothesis 10 indicates that more public items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB. The correlation between Public/Private and the other item-overall correlations was significant ( $r = .57, p < .05$ ). This means that the more public an item, the more other-raters use it in arriving at the overall CWB item. The correlation between Public/Private and self-rater item-overall correlations was not significant ( $r = .20, p > .10$ ). Hypothesis 10 was supported.

Hypothesis 11 indicated that more memorable items are expected to influence the other-rating of overall CWB rating more than the self-rating of overall CWB. The correlation between Memorability and other item-overall correlations was not significant ( $r = -.19, p > .10$ ). The correlation between Memorability and self-rater item-overall correlations was sizeable and significant ( $r = -.51, p < .05$ ). This indicates that the more

unmemorable an item, the more self-raters used it in the overall CWB rating.

Hypothesis 11 was not supported.

## 4. SUMMARY AND CONCLUSIONS

### 4.1 Summary

Overall, the results regarding the self-other correlation and mean differences hypotheses have several concluding points. First, the most important dimensions for predicting self-other agreement (i.e., the dimensions that displayed sizeable correlations with both self-other correlations and mean differences) are Observability, Public/Private, Task Relevance, and Memorability. Therefore, CWBs that are more observable, more public, more task unrelated, and more memorable display higher levels of self-other agreement. This also suggests that the reasons for disagreement between self- and other-raters of CWB are not due purely to idiosyncratic error of measurement. Rather, the CWBs about which self- and other-raters are most likely to disagree are predictable: they are those that are done in private settings and are difficult to observe, that are less task-related and therefore ambiguous, and are less memorable. The dimensions that predict disagreement all point toward issues with other-ratings rather than with self-ratings. That is, self-ratings should not be particularly affected by the observability, public/private setting, ambiguity (task-relatedness), or memorability of CWBs because the self-rater actually engaged in the CWBs and was at least somewhat aware of the intentions behind those actions. This leaves the other-raters as the source of the disagreement. Other-raters' ability to accurately capture employees' CWB will certainly be affected by factors such as the observability, ambiguity, and memorability of the CWBs. Berry et al. (2012) demonstrated that there is substantial disagreement between self- and other-raters of CWB; the present study demonstrates that much of this disagreement is due to



deficiencies in other-ratings of CWB (i.e., lack of opportunity to observe employees and lack of ability to interpret or remember certain CWBs).

What does this mean for CWB measurement? If other-ratings are useful for only certain CWBs, it is important to identify those behaviors. That is, it would be important to know if other-ratings are practical for only a subset of CWBs. The Berry et al. (2012) meta-analysis suggests that there may be more agreement between sources about CWB-I. I believe the present study has illustrated the particular subset of CWBs that other-raters are suited to rate. Table 6 lists the Observability, Public/Private, Memorability, and Task Relevance dimension ratings for each of the 19 CWBs; ratings are bolded for items that are “high” on a given dimension (i.e., above the midpoint of the 1-7 dimension rating scale). CWBs that are Observable, Public, Memorable and Task-Unrelated correspond to item 1 through item 7 on the Bennett and Robinson (2000) scale; these are the CWBs that other-raters have the greatest likelihood of being able to rate accurately. It is unlikely that other-raters can be trusted to rate employees’ accurately on the other CWBs (items 8-19). It is especially noteworthy that items 1-7 make up the entire CWB-I subscale, while items 8-19 make up the entire CWB-O scale. It seems that the CWB-I and CWB-O distinction captures the same items as the dimension-level examination. Thus, other-ratings of CWB-O should be viewed with skepticism. Self-ratings of CWB are probably typically more appropriate than other-ratings of CWB. If other-ratings must be used, either on their own or as a supplement to self-ratings, then they should probably only be used for CWB-I. This is similar to the conclusion of Berry et al. (2012), but the present study demonstrates *why* there is greater

agreement for the CWB-I scale than the CWB-O scale. CWB-Is are more observable, public, memorable, and task-unrelated than CWB-Os.

These conclusions are further reinforced by the results of the analyses predicting which CWB items most contribute to overall ratings of CWB for self- versus other-raters. Those results suggested that other-raters rely more heavily on observable, public, and memorable items when appraising overall CWB of an employee. This means that overall ratings given by other-raters (e.g., supervisors or coworkers) are disproportionately (positively) influenced by observable, public, and memorable items. These results shed light on what information is and is not being captured in the other-raters' overall CWB rating. Information related to more observable, public, and memorable CWBs is driving the other-raters' overall rating. However, for self-raters these qualities do not play a strong role in driving the self-overall CWB rating. For self-raters, their overall ratings are disproportionately influenced by socially desirable CWB items. This latter finding highlights that self-ratings, of course, have their own flaws and are susceptible to socially desirable responding (although, Berry et al., 2012 demonstrated the social desirability has only a small effect on the relationship between self-ratings of CWB and other variables).

These results are made all the more convincing because they align with the predictions of RAM Theory (Funder, 1995). RAM Theory proposes that accurate judgment of personality occurs through the following path: relevance of behavioral cues, availability for observation, detection, and utilization of relevant behavioral cues. In addition, the more accessible (memorable) the cue, the more likely it will be recalled and

utilized correctly in judgment of personality. The present study applied this pathway to examine accurate judgment of CWB. The findings support the predictions of RAM Theory and also indicate that RAM Theory can be applied to the CWB domain. Observability and memorability were the only dimensions hypothesized to influence self-other agreement based on RAM Theory, and both were found to be important in influencing self-other agreement on CWB. Examining the findings through the lens of RAM Theory can be insightful. Instead of having a collection of dimensions, RAM pathway provides a context for a number of the important dimensions. Namely, these dimensions influence self-other agreement because they influence various parts of the pathway from cue relevance to cue utilization. For example, the observability dimension influences the ability of a cue to be able to be seen by the observer. A cue must be able to be seen for other-raters to have the opportunity to observe it. Also, the memorability of an item influences the utilization of the cue. If a cue cannot be recalled properly, then it is likely it will not be used accurately in CWB judgment. The findings of this study confirm both of these predictions of RAM theory.

These explanations for agreement/disagreement, and the applications of RAM theory, could be illuminating not only for CWB measurement but also for measurement of other constructs within I/O psychology (e.g., task performance, OCB). Observability seems like an important factor in multi-source ratings regardless of domain. For example, it is not hard to imagine that more observable OCB's are more likely to be agreed upon across sources than more unobservable OCB's. Memorability may also influence agreement between other-raters and self-raters for task performance in a

similar fashion to CWB. That is, it is possible that more agreement is expected when task items/behaviors are more memorable.

#### **4.2 Practical Implications**

These results have practical implications. First, because other-raters are suited to rate only those behaviors that are observable, public, memorable, and task-unrelated, results suggest that other-ratings should be used only if researchers are interested in indexing behaviors with these qualities only (corresponding to item 1 through item 7; or CWB-I). Second, other-ratings of CWB may be used in industry to complement self-ratings. Results from the present study suggest that other-ratings should only be used for those CWBs that are observable, public, memorable, and task-unrelated. In other words, when organizations are interested in the other-rating for CWB, a shorter form that includes appropriate CWB-I items (as opposed to the whole measure) should be given to other-raters.

Because the present study suggests that self-ratings are generally preferred to other-ratings, I offer some practical recommendations for how to use self-ratings. For research purposes, self-ratings should be the method of choice because there is less incentive to score highly (i.e., and thus be dishonest). It is important to take multiple measures to assure participants that their responses will be anonymous and confidential, because responses are more likely to be honest (Berry et al., 2012). Examples of such measures are explicitly telling participants that their response are confidential and will only be used for research purposes, using secret codes that only the participant knows for data matching, and collecting data online rather than via hard copies (Berry et al., 2012).

However, for administrative purposes in applied settings, self-raters may be dishonest due to the incentive(s) they have to score highly. It is appropriate to supplement self-ratings with other-ratings; however, other-raters should only rate employees for the items on which they are qualified (i.e., CWB-I). If both CWB-O and CWB-I are of interest for administrative purposes, it may be most fruitful to collect self-ratings of CWB-I and CWB-O and to use other-ratings of CWB-I as an accuracy check (i.e., disregard an employee's self-ratings of both CWB-I and CWB-O if the self-rating of CWB-I is significantly lower than the other-rating of CWB-I). For developmental purposes (i.e., as opposed to award promotions, raises, bonuses etc.), use of other-ratings of CWB-O are likely to be useful feedback for employees. Even if the ratings of CWB-O from their supervisors and/or coworkers are not particularly accurate, employees would probably find it useful to know if those persons perceive them as engaging in high versus low levels of CWB-O.

#### **4.3 Limitations and Directions for Future Research**

The present study has some limitations. First, the main analyses in the present study used bivariate correlations and correlations do not allow researchers to examine unique effects of each dimension while holding other dimensions constant, as in multiple regression analysis. Unfortunately, multiple regression was not an option for the present study because the sample size was too small (i.e., 85) for the number of predictors (i.e., 19 items/predictors). For some regression analyses, the sample size would have only been 19 (e.g., regressing the CWB items on the 10 dimension ratings). Therefore, I was limited in the analyses available to examine the data. A limitation specifically for the

analyses predicting the overall ratings was that there were not enough participants per predictor (19 predictors). Ideally, enough data should be collected for a multiple regression to be stable enough to run. However, due to the dyad nature of the data (i.e., self-supervisor or self-coworker pairs), practical concerns for data collection included timeliness and cost-effectiveness.

Future research should apply the dimension-level examination of self-other agreement to other domains such as task performance or OCB. This and similar kinds of research endeavors for other measures can help explain why sources agree or disagree on the same measure. For example, a recent meta-analysis indicates that the self-other correlation of OCB ratings is far from perfect ( $\rho = .26$ ; Carpenter, Berry, & Houston, 2014). Conway and Huffcutt (1997) also found that self-other correlations of job performance ratings ranged from .14 to .22. It would be useful to apply the current paradigm to find out why sources differ on these kinds of ratings. Ultimately, this kind of research fuels a more precise understanding of ratings from different sources.

#### **4.4 Conclusions**

The goal of the present study was to leverage item-level analyses to advance understanding of why self- and other-raters disagree on employees' CWB. The findings of the present study indicate the most important dimensions for predicting self-other agreement are Observability, Public/Private, Task Relevance, and Memorability. Therefore, CWBs that are more observable, more public, more task-unrelated, and more memorable display higher levels of self-other agreement; CWB-Is, in particular, are characterized by these attributes. This suggests other-ratings are likely only appropriate

for this subset of CWBs and that other-ratings of CWB-O should be at the very least be viewed with skepticism.

## REFERENCES

- Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of applied psychology, 85*, 349-360.
- Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology, 97*, 613-636.
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: a review and meta-analysis. *Journal of Applied Psychology, 92*, 410-424.
- Bowling, N. A., & Gruys, M. L. (2010). Overlooked issues in the conceptualization and measurement of counterproductive work behavior. *Human Resource Management Review, 20*, 54-61.
- Budd, J. W., Arvey, R. D., & Lawless, P. (1996). Correlates and consequences of workplace violence. *Journal of Occupational Health Psychology, 1*, 197-210.
- Camara, W. J., & Schneider, D. L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist, 49*, 112-119.
- Carpenter, N. C., Berry, C. M., & Houston, L. (2014). A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior. *Journal of Organizational Behavior, 35*, 547-574.
- Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155-159.



- Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 331-360.
- Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of applied psychology, 90*, 1241-1255.
- Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychological Review, 102*, 652-670.
- Gruys, M. L., & Sackett, P. R. (2003). Investigating the dimensionality of counterproductive work behavior. *International journal of selection and assessment, 11*, 30-42.
- Harper, D. (1990). Spotlight abuse-save profits. *Industrial Distribution, 79*, 47-51.
- Hershcovis, M. S., Turner, N., Barling, J., Arnold, K. A., Dupre, K. E., Inness, M., . . . Sivanathan, N. (2007). Predicting workplace aggression: A meta-analysis. *Journal of Applied Psychology, 92*, 228–238. doi: 10.1037/0021-9010.92.1.228
- Oh, I. S., & Berry, C. M. (2009). The five-factor model of personality and managerial performance: validity gains through the use of 360 degree performance ratings. *Journal of Applied Psychology, 94*, 1498-1513.
- Oh, I. S., Charlier, S. D., Mount, M. K., & Berry, C. M. (2014). The two faces of high self-monitors: Chameleonic moderating effects of self-monitoring on the relationships between personality traits and counterproductive work behaviors. *Journal of Organizational Behavior, 35*, 92-111.

- Penney, L. M., & Spector, P. E. (2005). Job stress, incivility, and counterproductive work behavior (CWB): The moderating role of negative affectivity. *Journal of Organizational Behavior*, 26, 777-796.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88, 879-903.
- Sackett, P. R., **Berry, C. M.**, Wiemann, S., and Laczko, R. M. (2006). Citizenship and counterproductive work behavior: Clarifying relationships between the two domains. *Human Performance*, 19, 441-464.
- Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviors created equal?. *Journal of vocational behavior*, 68, 446-460.
- Tetlock, P. E., & Manstead, A. S. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy?. *Psychological Review*, 92, 59-77.
- Wallace, J. C. (2004). *The StudyResponse project: A Description and Evaluation of Using Standing Panels of Participants for Psychological Research* (Study Response Technical Report Number 13006) Syracuse: Syracuse University.

## APPENDIX A

### SME Pilot Test: Categories for CWB Items

1.) Observability: How observable is this CWB?

1= Not at all observable, 2, 3, 4= Somewhat observable, 5, 6, 7= Very observable

2.) Public/Private: How often would this CWB occur in a public (i.e., around at least one other person) vs. a private (i.e., the employee alone) setting?

1= Always in a private setting, 2= Usually in a private setting, 3= More often than not in a private setting, 4= Equally often in a public setting or private setting, 5= More often than not in a public setting, 6= Usually in a public setting, 7= Always in a public setting

3.) Organizational Sanctions: How strong is the punishment for this CWB likely to be if the employee is caught?

1= No punishment, 2= Lenient, 3= Somewhat lenient, 4= Neither strong nor lenient, 5= Somewhat strong, 6= Strong, 7= Very strong (e.g., termination or suspension)

4.) Base Rates: How often does the average employee engage in this CWB?

1= Never, 2= Once a year, 3= Twice a year, 4= Several times a year, 5= Monthly, 6= Weekly, 7= Daily

5.) Stigma: How stigmatized is this CWB? (Definition: A stigma is a mark of disgrace or a negative, social perception attached to an object)

1= Not at all stigmatized, 2, 3, 4= Somewhat stigmatized, 5, 6, 7= Very stigmatized

6.) Memorability: If an observer saw an employee engaging in this CWB, how memorable would this be to the observer?

1= Very unmemorable, 2= Unmemorable, 3= Somewhat unmemorable, 4= Neither memorable or unmemorable, 5= Somewhat memorable, 6= Memorable, 7= Very memorable

7.) Severity: How severe (i.e., minor acts versus more serious/severe deviance) is engagement in this CWB?

1= Not at all severe, 2, 3, 4= Somewhat severe, 5, 6, 7= Very severe

8.) Social Desirability: How socially desirable is it to engage in this CWB?

1= Very socially undesirable, 2= Socially undesirable, 3= Somewhat socially undesirable, 4= Neither desirable or undesirable, 5= Somewhat socially desirable, 6= Socially desirable, 7= Very socially desirable

9.) Ambiguity: How clear/ambiguous would it be that this behavior is a voluntary CWB?  
1= Very ambiguous, 2= Ambiguous, 3= Somewhat ambiguous, 4=Neither ambiguous nor clear, 5= Somewhat clear, 6= Clear, 7= Very clear

10.) Task Relevance: How related or unrelated is this CWB to job tasks?  
1= Completely unrelated, 2= Usually unrelated, 3= Somewhat unrelated, 4= Neither related nor unrelated, 5= Somewhat related, 6= Usually related, 7= Completely related

## APPENDIX B

### TABLES

**Table 1** Results of Study 1a: Dimension Scores for Each CWB Item

	Observ.	Public	Org. S.	Base R.	Stigma	Memor.	Severity	Social D.	Ambig.	Task R.
1. Made fun of someone at work	6.00	5.22	2.33	5.11	3.22	5.33	3.78	4.33	5.89	1.78
	(1.22)	(2.05)	(1.32)	(.78)	(1.48)	(1.32)	(.97)	(1.41)	(1.27)	(.97)
2. Said something hurtful to someone at work	5.78	5.33	2.44	4.44	4.00	5.33	4.56	3.00	5.89	1.89
	(1.20)	(1.58)	(1.13)	(.73)	(1.66)	(1.22)	(1.42)	(.71)	(.78)	(.93)
3. Made an ethnic, religious, or racial remark at work	5.22	4.33	5.00	2.67	6.44	6.78	6.67	1.44	6.56	1.44
	(1.56)	(1.66)	(1.41)	(1.50)	(.88)	(.44)	(.71)	(.73)	(.53)	(.53)
4. Cursed at someone at work	5.56	4.89	3.78	3.56	4.22	6.00	5.78	2.56	6.11	1.67
	(1.51)	(1.54)	(1.79)	(1.42)	(1.72)	(1.41)	(1.09)	(1.13)	(1.05)	(.71)
5. Played a mean prank on someone at work	5.67	5.11	3.78	2.56	4.11	6.56	5.11	3.56	6.44	1.56
	(1.50)	(1.36)	(.97)	(1.59)	(1.36)	(.53)	(.93)	(1.42)	(.73)	(.53)
6. Acted rudely toward someone at work	4.22	5.33	2.33	5.67	3.78	4.56	4.22	3.44	4.89	2.44
	(1.39)	(1.22)	(1.50)	(1.22)	(1.92)	(1.59)	(1.20)	(1.13)	(1.83)	(1.13)
7. Publicly embarrassed someone at work	6.44	6.78	3.11	3.44	5.11	6.50	5.78	2.56	6.22	1.78
	(1.13)	(.44)	(1.76)	(1.67)	(1.62)	(.76)	(.83)	(.88)	(1.09)	(.67)
8. Taken property from work without permission	2.13	1.78	5.22	3.33	4.56	5.78	5.56	2.44	6.33	3.56
	(.83)	(.44)	(1.30)	(1.22)	(1.74)	(1.09)	(1.67)	(1.01)	(.71)	(1.81)
9. Spent too much time fantasizing or daydreaming instead of working	2.56	1.78	3.00	6.22	1.67	2.89	2.78	3.33	3.78	5.44
	(1.59)	(.44)	(1.73)	(1.09)	(.87)	(1.69)	(1.72)	(.87)	(2.05)	(1.74)
10. Falsified a receipt to get reimbursed for more money than you spent on business expenses	2.13	1.56	6.33	2.00	5.11	6.67	6.78	1.78	6.67	3.33
	(1.13)	(.53)	(.87)	(1.00)	(2.09)	(.50)	(.44)	(.83)	(.71)	(1.87)
11. Taken an additional or longer break than is acceptable at your workplace	3.33	2.78	3.22	4.89	2.22	3.44	3.00	3.44	3.67	5.75
	(.71)	(.83)	(1.09)	(1.27)	(1.09)	(1.94)	(1.41)	(1.01)	(2.12)	(1.04)

12. Come in late to work without permission	4.67	3.44	3.33	5.00	2.63	3.56	3.22	3.44	3.22	5.75
	(1.12)	(1.59)	(1.12)	(.87)	(1.60)	(1.13)	(1.09)	(.88)	(1.99)	(1.16)
13. Littered your work environment	5.67	3.67	2.33	4.44	3.00	3.89	2.78	2.89	4.22	3.33
	(.87)	(1.58)	(1.22)	(1.94)	(2.00)	(1.83)	(1.39)	(1.05)	(1.72)	(1.66)
14. Neglected to follow your boss's instructions	4.67	3.00	5.22	4.11	3.75	4.78	4.89	2.89	3.67	6.78
	(.71)	(1.32)	(.67)	(1.76)	(1.98)	(.67)	(1.27)	(.93)	(1.50)	(.44)
15. Intentionally worked slower than you could have worked	3.56	2.44	3.44	5.00	2.89	3.67	3.78	3.11	3.22	6.67
	(1.33)	(.53)	(1.24)	(1.32)	(1.27)	(1.41)	(1.20)	(.93)	(2.05)	(.50)
16. Discussed confidential company information with an unauthorized person	3.22	3.67	6.67	1.78	5.00	6.44	6.67	1.67	5.56	4.11
	(1.48)	(2.24)	(.71)	(.67)	(2.18)	(.73)	(.71)	(.87)	(1.74)	(1.76)
17. Used an illegal drug or consumed alcohol on the job	4.56	2.33	6.56	1.78	6.75	6.89	6.89	2.11	6.00	5.11
	(1.24)	(.87)	(1.01)	(.97)	(.71)	(.33)	(.33)	(1.27)	(1.32)	(1.27)
18. Put little effort into your work	3.67	3.22	3.33	5.00	3.00	3.67	3.44	3.56	3.33	6.78
	(1.00)	(1.09)	(1.22)	(1.41)	(1.22)	(1.12)	(1.13)	(1.24)	(1.87)	(.44)
19. Dragged out work in order to get overtime	3.33	2.78	3.78	3.67	3.44	4.00	3.89	3.11	3.78	6.56
	(1.50)	(1.09)	(1.30)	(1.73)	(1.24)	(1.12)	(1.05)	(.93)	(1.92)	(.53)

*Note.* Each CWB item has two corresponding rows. The top row indicates the score for that dimension, which is the mean. The bottom row indicates the SD, which is in parentheses.

**Table 2** Intercorrelation Matrix: Dimensions

	1	2	3	4	5	6	7	8	9
1. Observability									
2. Public/Private	.843**								
3. Organizational Sanctions	-.429	-.477*							
4. Base Rates	-.011	-.050	.816**						
5. Stigma	.174	.159	.674**	.842**					
6. Memorability	.267	.311	.593**	.853**	.907**				
7. Severity	.015	.078	.785**	.880**	.933**	.932**			
8. Social Desirability	-.200	-.198	.772**	.764**	.770**	.633**	.808**		
9. Ambiguity	-.257	-.357	-.360	-.669**	-.760**	-.917**	-.782**	-.493*	
10. Task Relevance	-.553*	-.691**	.146	-.272	-.437	-.645**	-.412	-.176	.832**

*Note.* \*\*. Correlation is significant at the .01 level (two-tailed). \*. Correlation is significant at the .05 level (two-tailed).

**Table 3** Intercorrelation Matrix: Study Variables

	Mean	SD	Sex	Age	Race	CWB1	CWB2	CWB3	CWB4	CWB5	CWB6	CWB7	CWB8	CWB9	CWB10
<b>Mean</b>	-	-	-	-	-	2.21	1.88	1.71	1.86	1.66	1.84	1.63	1.33	2.17	1.34
<b>SD</b>	-	-	-	-	-	1.95	1.69	1.67	1.83	1.48	1.67	1.41	1.06	1.81	1.15
<b>Sex</b>	-	-	-	.219*	-.027	-.137	.046	.010	.029	.050	.068	.057	.085	-.032	.070
<b>Age</b>	-	-	-.023	-	.113	-.297**	-.210*	-.247*	-.143	-.197	-.194	-.223*	.023	-.338**	-.014
<b>Race</b>	-	-	-.012	.171	-	.029	-.021	-.076	-.100	-.158	-.101	-.124	-.109	.061	-.129
<b>CWB1</b>	2.63	2.09	-.087	-.184	.067	<b>.476**</b>	.676**	.721**	.555**	.603**	.631**	.639**	.401**	.676**	.378**
<b>CWB2</b>	1.86	1.51	-.030	-.180	.067	.474**	<b>.589**</b>	.848**	.889**	.845**	.928**	.840**	.464**	.794**	.563**
<b>CWB3</b>	1.86	1.79	-.037	-.162	.037	.463**	.637**	<b>.545**</b>	.787**	.827**	.806**	.859**	.519**	.768**	.610**
<b>CWB4</b>	1.95	1.74	-.214*	-.141	.090	.555**	.617**	.676**	<b>.491**</b>	.891**	.900**	.818**	.555**	.659**	.615**
<b>CWB5</b>	1.76	1.58	-.092	-.123	.076	.411**	.689**	.771**	.803**	<b>.516**</b>	.865**	.928**	.649**	.711**	.721**
<b>CWB6</b>	1.66	1.38	-.064	-.178	.001	.408**	.830**	.529**	.654**	.602**	<b>.380**</b>	.901**	.495**	.737**	.557**
<b>CWB7</b>	1.39	1.14	-.043	-.209*	.068	.241*	.749**	.561**	.630**	.608**	.793**	<b>.397**</b>	.597**	.736**	.686**
<b>CWB8</b>	1.59	1.19	.205*	-.153	.023	.189	.490**	.439**	.300**	.413**	.546**	.492**	<b>.237**</b>	.424**	.857**
<b>CWB9</b>	2.77	1.88	.074	-.247*	.054	.556**	.516**	.499**	.457**	.437**	.489**	.356**	.415**	<b>.377**</b>	.535**
<b>CWB10</b>	1.32	0.96	-.014	.071	-.006	.173	.337**	.340**	.208*	.290**	.303**	.207*	.493**	.227*	<b>.555**</b>
<b>CWB11</b>	2.72	1.91	.091	-.185	.023	.419**	.481**	.506**	.415**	.448**	.434**	.260*	.509**	.711**	.266**
<b>CWB12</b>	2.09	1.62	-.001	-.192	-.154	.516**	.271**	.396**	.453**	.335**	.392**	.185	.242*	.582**	.196
<b>CWB13</b>	1.76	1.36	.126	.065	.255*	.062	.356**	.214*	.331**	.461**	.390**	.385**	.414**	.348**	.147
<b>CWB14</b>	1.86	1.37	.059	-.241*	-.039	.367**	.360**	.317**	.409**	.321**	.519**	.311**	.482**	.560**	.071
<b>CWB15</b>	2.15	1.79	-.039	-.323**	-.054	.553**	.541**	.494**	.574**	.436**	.638**	.439**	.448**	.724**	.090
<b>CWB16</b>	1.53	1.19	-.112	-.032	.037	.534**	.412**	.463**	.427**	.307**	.399**	.336**	.407**	.410**	.336**
<b>CWB17</b>	1.56	1.44	.046	-.140	.227*	-.009	.230*	.129	.307**	.361**	.309**	.392**	.250*	.142	.116
<b>CWB18</b>	2.14	1.63	.054	-.190	.068	.434**	.528**	.504**	.387**	.350**	.497**	.491**	.565**	.702**	.225*
<b>CWB19</b>	1.54	1.34	-.108	-.034	-.138	.408**	.351**	.456**	.398**	.405**	.384**	.245*	.198	.399**	.378**
<b>CWB20</b>	2.46	1.79	.025	-.118	.111	.552**	.606**	.629**	.590**	.508**	.543**	.430**	.451**	.665**	.262*



**Table 3** (continued)

	CWB11	CWB12	CWB13	CWB14	CWB15	CWB16	CWB17	CWB18	CWB19	CWB20
<b>Mean</b>	1.88	1.91	1.84	1.87	1.77	1.38	1.49	1.65	1.44	2.11
<b>SD</b>	1.54	1.50	1.76	1.66	1.60	1.20	1.36	1.35	1.16	1.61
<b>Sex</b>	-.019	-.103	-.045	.078	.073	.104	-.044	.046	.135	.010
<b>Age</b>	-.170	-.287**	-.214*	-.173	-.250*	-.087	-.111	-.282**	-.149	-.285**
<b>Race</b>	.012	-.017	.069	-.031	-.050	-.090	.006	-.084	.017	-.081
<b>CWB1</b>	.539**	.626**	.586**	.562**	.608**	.447**	.483**	.546**	.334**	.731**
<b>CWB2</b>	.750**	.712**	.776**	.840**	.725**	.679**	.571**	.764**	.536**	.793**
<b>CWB3</b>	.730**	.799**	.830**	.782**	.754**	.670**	.637**	.789**	.572**	.821**
<b>CWB4</b>	.705**	.664**	.711**	.842**	.718**	.641**	.613**	.737**	.548**	.671**
<b>CWB5</b>	.664**	.682**	.751**	.855**	.789**	.691**	.653**	.790**	.610**	.756**
<b>CWB6</b>	.687**	.623**	.698**	.848**	.703**	.645**	.561**	.765**	.530**	.766**
<b>CWB7</b>	.669**	.673**	.748**	.851**	.789**	.705**	.647**	.861**	.607**	.804**
<b>CWB8</b>	.506**	.416**	.502**	.462**	.561**	.679**	.760**	.521**	.515**	.460**
<b>CWB9</b>	.643**	.691**	.822**	.703**	.672**	.635**	.645**	.717**	.641**	.788**
<b>CWB10</b>	.589**	.545**	.554**	.553**	.664**	.774**	.746**	.628**	.559**	.571**
<b>CWB11</b>	<b>.494**</b>	.811**	.692**	.734**	.639**	.513**	.583**	.626**	.410**	.692**
<b>CWB12</b>	.595**	<b>.494**</b>	.677**	.738**	.688**	.490**	.539**	.628**	.389**	.719**
<b>CWB13</b>	.487**	.282**	<b>.490**</b>	.713**	.705**	.654**	.729**	.682**	.718**	.674**
<b>CWB14</b>	.631**	.661**	.404**	<b>.460**</b>	.777**	.621**	.585**	.802**	.564**	.732**
<b>CWB15</b>	.695**	.653**	.292**	.720**	<b>.270**</b>	.748**	.613**	.781**	.627**	.778**
<b>CWB16</b>	.511**	.318**	.178	.265**	.509**	<b>.274**</b>	.631**	.683**	.699**	.674**
<b>CWB17</b>	.209*	.105	.569**	.222*	.129	.050	<b>.460**</b>	.668**	.768**	.514**
<b>CWB18</b>	.617**	.489**	.283**	.568**	.714**	.563**	.143	<b>.296**</b>	.647**	.729**
<b>CWB19</b>	.522**	.466**	.084	.334**	.509**	.459**	.019	.465**	<b>.153</b>	.524**
<b>CWB20</b>	.679**	.522**	.323**	.516**	.640**	.416**	.084	.616**	.471**	<b>.541**</b>

*Note.* \*. Correlation is significant at the 0.05 level (2-tailed). \*\*. Correlation is significant at the 0.01 level (2-tailed). The diagonal represents the self-other correlation for the specified pair of items. The other-rater correlations are above the diagonal. The self-rater correlations are below the diagonal.

**Table 4** Summary of Hypotheses, Research Questions, and Their Results

<b>Hypothesis/Research Question</b>	<b>Statistic</b>	<b>Significant?</b>	<b>Interpretation</b>
<i>Self-Other Correlations</i>			
(1a) Self-other correlations will be stronger for more observable CWBs	$r = .51$	yes	Hypothesis 1a was supported, meaning the more observable the CWB, the stronger the self-other correlations
(2a) Self-other correlations will be stronger for more public CWBs	$r = .31$	no	Hypothesis 2a was not supported
(3a) Self-other correlations will be stronger for more memorable CWBs	$r = .29$	no	Hypothesis 3a was not supported
(4a) Self-other correlations will be stronger for more unambiguous (clear) CWBs	$r = .36$	no	Hypothesis 4a was not supported
(5a) Self-other correlations will be stronger for more socially desirable CWBs	$r = -.05$	no	Hypothesis 5a was not supported
(RQ1a) How does Severity influence self-other correlations of CWBs	$r = .14$	no	No relationship was found
(RQ2a) How does Task Relevance influence self-other correlations of CWBs	$r = -.50$	yes	This means the more task-unrelated the CWB, the stronger the self-other correlations
(RQ3a) How does Base Rates influence self-other correlations of CWBs	$r = .09$	no	No relationship was found
(RQ4a) How does Organizational Sanctions influence self-other correlations of CWBs	$r = .09$	no	No relationship was found

(RQ5a) How does Stigma influence self-other correlations of CWBs	$r = .17$	no	No relationship was found
<i>Mean Differences</i>			
(1b) Mean differences will be smaller for more observable CWBs	$r = -.50$	yes	Hypothesis 1b was supported, meaning the more observable the CWB, the smaller the mean differences
(2b) Mean differences will be smaller for more public CWBs	$r = -.51$	yes	Hypothesis 2b was supported, meaning the more public the CWB, the smaller the mean differences
(3b) Mean differences will be smaller for more memorable CWBs	$r = -.49$	yes	Hypothesis 3b was supported, meaning the more memorable the CWB, the smaller the mean differences
(4b) Mean differences will be smaller for more unambiguous (clear) CWBs	$r = -.42$	yes	Hypothesis 4b was supported, meaning the less ambiguous a CWB, the smaller the mean differences
(5b) Mean differences will be smaller for more socially desirable CWBs	$r = .32$	no	Hypothesis 5b was not supported
(RQ1b) How does Severity influence self-other mean differences of CWBs	$r = -.42$	yes	This means the more severe the CWB, the smaller the mean differences
(RQ2b) How does Task Relevance influence self-other mean differences of CWBs	$r = .48$	yes	This means the more task-unrelated the CWB, the less mean differences
(RQ3b) How does Base Rates influence self-other mean differences of CWBs	$r = .33$	no	No relationship was found

(RQ4b) How does Organizational Sanctions influence self-other mean differences of CWBs	$r = -.05$		no		No relationship was found
(RQ5b) How does Stigma influence self-other mean differences of CWBs	$r = -.50$		yes		The more stigma attached to the behavior, the smaller the mean differences
<i>Item-overall Correlation (Self vs. Other)</i>	<b>Self Statistic</b>	<b>Significant?</b>	<b>Other Statistic</b>	<b>Significant?</b>	<b>Interpretation</b>
(6) More socially undesirable items are expected to influence the overall other rating more than the overall self-rating	$r = .46$	yes	$r = .28$	no	Hypothesis 6 was not supported
(7) More unambiguous items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB	$r = -.37$	no	$r = -.16$	no	Hypothesis 7 was not supported
(8) More task-unrelated items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB	$r = -.05$	no	$r = .24$	no	Hypothesis 8 was not supported
(9) More observable items are expected to influence the other-rating of overall CWB rating more than the self-rating of overall CWB rating	$r = .03$	no	$r = .50$	yes	Hypothesis 9 was supported, meaning the more observable the CWB, the more it was used in the other overall rating compared to the self overall rating
(10) More public items are expected to influence the other-rating of overall CWB more than the self-rating of overall CWB	$r = .20$	no	$r = .57$	yes	Hypothesis 10 was supported, meaning the more public the CWB, the more it was used in the other overall rating compared to the self overall rating

(11) More memorable items are expected to influence the other-rating of overall CWB rating more than the self-rating of overall CWB	$r = -.19$	no	$r = -.51$	yes	Hypothesis 11 was not supported
---	------------	----	------------	-----	---------------------------------

**Table 5** Cohen's D for Each CWB Item

<b>Item</b>	<b>Cohen's <i>d</i></b>
1. Made fun of someone at work	0.21
2. Said something hurtful to someone at work	-0.01
3. Made an ethnic, religious, or racial remark at work	0.09
4. Cursed at someone at work	0.05
5. Played a mean prank on someone at work	0.06
6. Acted rudely toward someone at work	-0.12
7. Publicly embarrassed someone at work	-0.19
8. Taken property from work without permission	0.23
9. Spent too much time fantasizing or daydreaming instead of working	0.33*
10. Falsified a receipt to get reimbursed for more money than you spent on business expenses	-0.02
11. Taken an additional or longer break than is acceptable at your workplace	0.48*
12. Come in late to work without permission	0.12
13. Littered your work environment	-0.05
14. Neglected to follow your boss's instructions	-0.01
15. Intentionally worked slower than you could have worked	0.23
16. Discussed confidential company information with an unauthorized person	0.12
17. Used an illegal drug or consumed alcohol on the job	0.05
18. Put little effort into your work	0.32*
19. Dragged out work in order to get overtime	0.08
20. Overall, how often do you engage in counterproductive work behaviors?	0.21

*Note.* \*. There is a statistical difference between means at the .05 level (two-tailed). Positive *d*-values indicate that the self-rating is higher than the other-rating.

**Table 6** Qualities that Characterize Each CWB Item

	Observability	Public/Private	Memorability	Task Relevance	Total
1. Made fun of someone at work	<b>6.00</b>	<b>5.22</b>	<b>5.33</b>	<b>6.22</b>	4
2. Said something hurtful to someone at work	<b>5.78</b>	<b>5.33</b>	<b>5.33</b>	<b>6.11</b>	4
3. Made an ethnic, religious, or racial remark at work	<b>5.22</b>	<b>4.33</b>	<b>6.78</b>	<b>6.56</b>	4
4. Cursed at someone at work	<b>5.56</b>	<b>4.89</b>	<b>6.00</b>	<b>6.33</b>	4
5. Played a mean prank on someone at work	<b>5.67</b>	<b>5.11</b>	<b>6.56</b>	<b>6.44</b>	4
6. Acted rudely toward someone at work	<b>4.22</b>	<b>5.33</b>	<b>4.56</b>	<b>5.56</b>	4
7. Publicly embarrassed someone at work	<b>6.44</b>	<b>6.78</b>	<b>6.50</b>	<b>6.22</b>	4
8. Taken property from work without permission	2.13	1.78	<b>5.78</b>	<b>4.44</b>	2
9. Spent too much time fantasizing or daydreaming instead of working	2.56	1.78	2.89	2.56	0
10. Falsified a receipt to get reimbursed for more money than you spent on business expenses	2.13	1.56	<b>6.67</b>	<b>4.67</b>	2
11. Taken an additional or longer break than is acceptable at your workplace	3.33	2.78	3.44	2.25	0
12. Come in late to work without permission	<b>4.67</b>	3.44	3.56	2.25	1
13. Littered your work environment	<b>5.67</b>	3.67	3.89	<b>4.67</b>	2
14. Neglected to follow your boss's instructions	<b>4.67</b>	3.00	<b>4.78</b>	1.22	2
15. Intentionally worked slower than you could have worked	3.56	2.44	3.67	1.33	0
16. Discussed confidential company information with an unauthorized person	3.22	3.67	<b>6.44</b>	3.89	1
17. Used an illegal drug or consumed alcohol on the job	<b>4.56</b>	2.33	<b>6.89</b>	2.89	2
18. Put little effort into your work	3.67	3.22	3.67	1.22	0
19. Dragged out work in order to get overtime	3.33	2.78	4.00	1.44	0

*Note.* This table indicates the total number of qualities (i.e., that are prone to highest self-other agreement) that characterize each CWB item. For all 4 dimensions/qualities, ratings greater than a 4.00 indicate that the item is characterized by that quality.