Spring 2022

Does Insufficient Effort Responding Account for the Relationship Between Psychopathy and Social Desirability?

Logan F. Folger

DOES INSUFFICIENT EFFORT RESPONDING ACCOUNT FOR THE RELATIONSHIP BETWEEN PSYCHOPATHY AND SOCIAL DESIRABILITY?

by

LOGAN FOLGER

(Under the Direction of Nicholas Holtzman)

ABSTRACT

This research investigated the relationship between self-report psychopathy and social desirability and whether insufficient effort responding may impact this relationship. It was hypothesized that insufficient effort responding is biasing the relationship. Therefore, the significant negative relationship between psychopathy and social desirability may approach zero after isolating insufficient effort responding data. Two studies were conducted to aid this investigation. The first study utilized the Monte Carlo method to simulate this relationship. The simulation had 1,000 trials using 1,000 participants with proportions of valid respondents ranging from 0.0 to 1.0 with increments of 0.1. The results showed that insufficient effort responding may be biasing this relationship between psychopathy and social desirability, so an investigation among human participants was warranted. The second study examined this relationship among human participants using a community sample. The sample included 300 participants, and each participant was asked to take the Self-report Psychopathy scale and the PRF Desirability subscale. Insufficient effort responding was isolated using response time, odd-even consistency, long-string analysis, and Mahalanobis distance. The results showed that after isolating insufficient effort responding data, the relationship between psychopathy and social desirability became non-significant but not for the reasons predicted. The insufficient effort responding data did not surround the midpoint as expected; rather, the valid data surrounded the

midpoint. Additionally, the TOST procedure was done to test for statistical equivalence and found this relationship may be significant. Even though the simulation showed the possibility of a biased relationship, the results of study 2 produced a contradicting conclusion.
INDEX WORDS: Insufficient effort responding, Psychopathy, Social desirability

DOES INSUFFICIENT EFFORT RESPONDING ACCOUNT FOR THE RELATIONSHIP BETWEEN PSYCHOPATHY AND SOCIAL DESIRABILITY?

by

LOGAN FOLGER

B.S., University of Georgia, 2020

B.A., University of Georgia, 2020

A Thesis Submitted to the Graduate Faculty of Georgia Southern University

in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

COLLEGE OF BEHAVIORAL AND SOCIAL SCIENCES

DOES INSUFFICIENT EFFORT RESPONDING ACCOUNT FOR THE RELATIONSHIP BETWEEN PSYCHOPATHY AND SOCIAL DESIRABILITY?

by

LOGAN FOLGER

Major Professor:	Nicholas Holtzman
Committee:	Amy Hackney
	Chad Posick

Electronic Version Approved:
May 2022

DEDICATION

This thesis is dedicated to my parents, James and Evelyn, who always encouraged me and supported me in my education. I would not be where I am without their support. Thank you for all that you have done for me.

ACKNOWLEDGMENTS

I want to thank my faculty advisor and thesis chair, Dr. Nicholas Holtzman, for his constant support and guidance throughout the development of this project and my time during the program. Thank you for challenging me, supporting me, and encouraging me. I would also like to thank my committee members, Dr. Amy Hackney and Dr. Chad Posick, for their time and helpful feedback in the review of my proposal and final draft. Thank you to my amazing friends, Olivia, Haley, and Grant, for your constant emotional support and encouragement, and my cohort, who made this graduate school journey so enjoyable. Lastly, I want to thank the Graduate Student Organization (GSO) for funding this project and making this research possible.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Psychopathy is one of the most studied personality disorders despite not being included in official diagnostic manuals (Lynam et al., 2011). A recent search on PsycInfo for peer-reviewed articles with the keyword *psychopathy* in the title or subject resulted in 5,995 articles. A search for *antisocial personality disorder* peer-reviewed articles resulted in 7,734. The most researched personality disorder was *borderline personality disorder* with a total result of 9,551 peer-reviewed articles on the topic. There are various conceptualizations of psychopathy that can be used to describe the psychopathic personality. Cleckley (1941) first conceptualized psychopathy in his book *The Mask of Sanity*. He described the psychopath as a superficially charming person that can make a strong positive impression. Additional to superficial charm is their lack of remorse, responsibility, and insight, and their inclination to engage in antisocial, manipulative, and deceitful behavior.

Derived from Cleckley's (1941) conceptualization was the Psychopathy Checklist Revised (PCL-R), and through factor analysis, two-factor and four-factor models emerged from the PCL-R (Hare, 1991, 2003; Vitacco et al., 2005). The two factors produced were interpersonal/affective and social deviation (Pérez et al., 2015). The four factors included interpersonal manipulation, callous affect, erratic lifestyle, and criminal tendencies (Gordts et al., 2017; Hare, 2003). The PCL-R is the most common and well-known diagnostic tool for psychopathy; however, it is most appropriate for use in institutionalized samples because it requires examination of criminal records, a detailed interview, and extensive case history (Hare, 1991). The conceptualization of psychopathy has produced several different factor models, but the frameworks all surround similar traits such as deceitfulness, interpersonal manipulation, egocentricity, etc. (Hare & Neumann, 2005; Neumann & Hare, 2008). Self-reported psychopathy measures have been

derived from these models as well (LSRP: Levenson et al., 1995; PPI-R: Lilienfeld & Widows, 2005; SRP-III: Williams et al., 2007). These measures are also related with correlations ranging from .61 to .80 (Wilson et al., 2011).

Although there are different factor models of psychopathy, these conceptualizations are based on common traits of psychopathy such as being manipulative, callous, egocentric and lacking remorse, empathy, and self-control (Miller et al., 2011; Neumann & Hare, 2008). With callous affect being a core trait of psychopathy, psychopaths' lack of remorse and self-control and their tendency to use others for their own gain may lead to corrupt and antisocial behaviors, but they may not necessarily be unlawful (Hare, 1991; Neumann et al., 2005; Neumann & Hare, 2008). Studies have shown that perspective taking can allow an individual high in psychopathy to experience empathy-related responses if they can make a conscious effort to see themselves in someone else's tragic situation (Beussink et al., 2017; Meffert et al., 2013). Self-reported measures of psychopathy were developed to measure the prevalence of psychopathy in the community/non-institutionalized samples. While researchers have tried to examine the prevalence of psychopathy in community samples using self-reported measures of psychopathy, it still remains difficult to estimate prevalence. Using the screening version of the PCL (PCL:SV), Neumann and Hare (2008) found 1-2% of their community sample as exhibiting potential psychopathic tendencies, and it is estimated that 15-25% of the general offender population may possess psychopathic traits (Lilienfeld & Arkowitz, 2007; Woodworth & Porter, 2002).

Psychopathy and Social Desirability

With defining characteristics such as manipulation, egocentricity, and deception, it can be assumed that psychopaths may try to act socially desirable or "faking good" in certain situations. Specifically, concerns about socially desirable responding on self-reported psychopathy measures

are warranted. Social desirability can be defined as one's tendency to describe themselves with seemingly socially desirable statements and reject socially undesirable statements about their personality (Edwards, 1958). Therefore, socially desirable responding is one type of response bias that warrants considerable concern on self-reported psychopathy measures, and many researchers have studied this type of response distortion on self-reported psychopathy measures along with malingering, or faking bad (Book et al., 2006; MacNeil & Holden, 2006; Ray et al., 2013; Watts et al., 2016).

It can be assumed that due to the deceptive and manipulative nature of psychopaths, socially desirable responding would present as a validity problem on self-reported psychopathy measures; however, the literature has actually found a negative relationship between social desirability and psychopathy (Kowalski et al., 2018; Ray et al., 2013; Verschuere et al., 2014). Contrary to this literature, researchers have also found that those with higher psychopathy scores were more likely to fake good, or answered in a socially desirable way, on psychological measures than those with lower psychopathy scores (Book et al., 2006; Fisher et al., 2018). Furthermore, when Ray and colleagues (2013) conducted a meta-analysis on self-reported psychopathy and distorted response styles, they found that factor II of the Psychopathic Personality Inventory-Revised (PPI-R) had a positive (though not significant) relationship to social desirability. This is consistent with previous research that found Machiavellian egocentricity, blame externalization, and stress immunity, which are subcomponents of factor II, to enable successful faking measured by multiple faking scales including the PRF Desirability scale (MacNeil & Holden, 2006). However, the overall conclusion from Ray and colleagues' (2013) investigation is that psychopathy was either negatively or negligibly associated with social desirability.

Given that the literature on this is mixed, there are several possible explanations. Because certain components of psychopathy were found to be positively associated with faking good, individuals that have extremely high levels of certain components of psychopathy may be more likely to exhibit faking behavior (MacNeil & Holden, 2006; Ray et al., 2013). The second explanation may be the nature of the study itself. The studies that found a positive relationship between psychopathy and social desirability included either instructions to fake or incentives to fake despite warnings about faking (Book et al., 2006; Fisher et al., 2018). Psychopaths may only act socially desirable when there are incentives and are better at faking because of their psychopathic traits, but this does not explain the negative relationship.

A possible explanation for the negative relationship between psychopathy and social desirability may be insufficient effort responding. The true relationship between psychopathy and social desirability may be negligible, but when motivated enough (e.g., by offering incentives), psychopathic individuals may exhibit faking. If psychopaths do engage in socially desirable responding, this would make the relationship between psychopathy and social desirability positive and may also have an implication for their true psychopathy scores.

Insufficient Effort Responding

There are many concerns that come with self-reported personality measures, whether it is due to socially desirable responding, malingering, lack of understanding, or random responding, and many researchers have tried to incorporate validity scales into their tests as a preventative measure (Curran, 2016; Groth & Kleka, 2018; Ray et al., 2013). Because personality psychology research is partially dependent on self-report data, it is important to have these preventative measures as well as analytic techniques that can help identify biased data (Curran, 2016). However, despite efforts to implement instructional manipulation checks and other preventative measures, some invalid data may still be misidentified as valid because instructional

manipulation checks may fail (Hauser & Schwarz, 2016). There are three classes of invalid data as identified by Johnson (2005). The first is linguistic incompetence or misunderstanding which concerns the issue of understanding and interpreting the testing items according to one's own linguistic knowledge. The second is misrepresentation such as misrepresenting one's self as more socially desirable than they actually are or presenting one's self as more sinister than they actually are. This type of responding is usually more common in high stake settings (Curran, 2016). The last class is carelessness and inattentiveness type of responding. This is more common in unsupervised research settings such as online surveys or class required participation. Carelessness and inattentiveness can occur when participants are unmotivated and are not thoughtfully reading and answering the questions (Curran, 2016).

For many years, inattentive or careless respondents were referred to as random respondents because it was thought that this type of random responding also produced random patterns of result and thus random error (Beach, 1989; Curran, 2016). However, this is not entirely the case as Huang and colleagues (2012) found that inattentive responding can endorse non-random patterns of responding as well; thus, the description of this type of responding style became known as careless (Meade & Craig, 2012) or insufficient effort responding (Huang et al., 2012). Because insufficient effort responding can produce both random and non-random patterns of responding, it can potentially be disguised as valid data making correlations more extreme (Holtzman & Donnellan, 2017; Huang et al., 2012). Although patterned responses can be expected from insufficient effort responding, it is thought that these responses will produce a mean in the midpoint of the scale after averaging the scores (Huang et al., 2012). There are three criteria that must be met to make correlations more extreme (Holtzman & Donnellan, 2017; Huang et al., 2012). The first criteria is met when there are valid responses from participants whose averages fall away from where the center point would be; these are referred to as valid

uncentered responses. The second criteria is met when careless responding yields mean responses that fall on the midpoint of the scale (for instance, a 3 on a 1-to-5 Likert scale); this is referred to as invalid centered responses. When invalid centered responses and valid uncentered responses occur for both of the variables that are entered into the correlation, they can create a seemingly linear relationship between the two variables where none exists (Holtzman & Donnellan, 2017; Huang et al., 2012). Figure 1 shows a visual representation of this scenario.

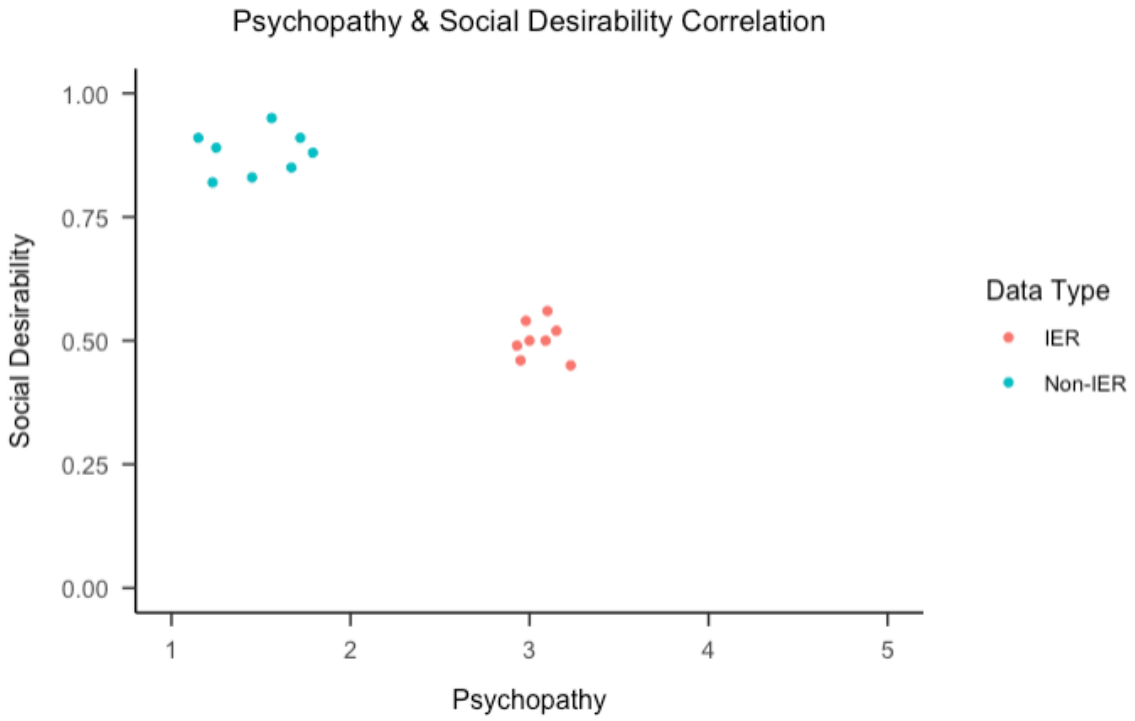

Figure 1. This graph depicts the correlation between psychopathy and social desirability with and without insufficient effort responding in the data. As shown in the graph, when insufficient effort responding is added to the data, the relationship between psychopathy and social desirability becomes more linear.

The blue dots in the scatterplot represent data without insufficient effort responding, and the red dots represent data with insufficient effort responding. As the scatterplot shows, the blue dots represent a weakly correlated relationship. However, if the careless respondents (red dots)

are added to the whole dataset, that can create a seemingly negative linear relationship between the two variables. Based on the simulation results and previous research, it is possible that insufficient effort responding has a greater impact on the relationship between self-reported psychopathy and social desirability than previously thought. Although researchers have tried to implement instructional manipulation checks into their surveys, they are not 100% preventative (Hauser & Schwarz, 2016). Additionally, some personality measures will contain validity scales as well though they do not always add to the validity of the measure (Piedmont et al., 2000). Validity scales and instructional manipulation checks may not be able to detect all careless and invalid respondents, and when 5% of careless respondents may contribute to producing significant results, it becomes increasingly concerning (Huang et al., 2015).

My project aims to examine whether insufficient effort responding accounts for the relationship between socially desirable responding and self-reported psychopathy. This investigation begins with a simulation of data using Holtzman and Donnellan's (2017) random respondent simulator that showed that insufficient effort responding could be the potential factor behind the significant negative relationship between psychopathy and social desirability. The second part of this study involves human participants and is in part a replication of Kowalski and colleagues' (2018) study on psychopathy and social desirability. Kowalski et al. (2018) examined the relationships between the dark triad and socially desirable responding and self-monitoring. This is one of the studies that found a significant negative relationship between psychopathy and social desirability. Although the authors noted that they had attention checks, previous research showed that attention checks may not always be effective (Curran & Hauser, 2015). Curran and Hauser (2015) found that participants may answer attention check questions inaccurately even after reading the items aloud. Although it was recorded in less than 10% of individuals, Curran (2016) recommends using different types of attention checks with caution. Additionally, this

study takes place on Amazon Mechanical Turk (MTurk), an online survey system. MTurk provided a suitable pool of participants for this study because it may be especially vulnerable to insufficient effort responding. MTurkers have been recorded to complete studies faster than other platform populations as well as engaging in multitasking while completing studies (Kees et al., 2017; Necka et al., 2016).

All techniques used to detect insufficient effort responding have a risk of including invalid respondents and excluding valid respondents. However, the combination of these techniques can help ensure that the responses included are the most valid. This study uses multiple techniques to detect insufficient effort responding with a specific focus on the relationship between self-reported psychopathy and social desirability to address the mixed results in the psychopathy literature. Through Holtzman and Donnellan's (2017) random respondent simulator, it is predicted that the negative relationship between psychopathy and socially desirable responding may approach zero after eliminating insufficient effort responding data. If this relationship does turn out to be attenuated, a Two One-Sided Tests (TOST procedure) will be carried out to test for statistical equivalence (Lakens, 2017). The TOST procedure is typically employed to test whether an effect or relationship is statistically equivalent to zero.

CHAPTER 2

STUDY 1

A simulation was conducted using Holtzman and Donnellan's (2017) random respondent simulator to examine the possibility of insufficient effort responding creating a seemingly greater association between psychopathy and social desirability than what actually exists. Because of the mixed results in previous research, the possibility of a confounding variable is very much in the question, especially in self-reported data. This simulation was conducted in order to determine whether a human study was warranted. If the simulated data is able to attenuate the correlation between psychopathy and social desirability depending on the proportion of insufficient effort responding, then that necessitates a reason to investigate this relationship further. Additionally, if the relationship weakened when insufficient effort responding was taken out of the data, then there would be a foundation to the investigation.

Method

For the Self-Reported Psychopathy scale, the following measurements were used: (1) the item number used was 64, (2) lowest response option was 1, (3) highest response option was 5, (4) the mean of valid responses was 2.11 based on Kowalski et al.'s (2018) data, (5) the standard deviation across participants was 0.39 based on Kowalski et al's (2018) data, (6) a within-person standard deviation of 1.43 was used based on Holtzman and Donnellan's (2017) reference data (Paulhus et al., 2016), and (7) the mean of invalid responses was 3. For the PRF Desirability scale, the following specifications were used: (1) the item number used was 16, (2) lowest response option was 0, (3) highest response option was 1, (4) the mean of valid responses was 0.76 based on Kowalski et al's (2018) data, (5) the standard deviation across participants was 0.15 based on Kowalski et al.'s (2018) data, (6) three estimated within-person standard deviations were used: 0.15, 0.20, and 0.25, and (7) the mean of invalid responses was 0.5. I used three

estimated of within-person standard deviation because to my knowledge, no data are available on this topic.

One thousand Monte Carlo simulations were run with proportions of valid respondents ranging from 0.0 to 1.0 with increments of 0.1. The Monte Carlo simulations were run based on these numbers for each of the different estimated within-person standard deviation for the PRF Desirability scale, and 1,000 participants were used in each simulation. In total, three 1,000 trials of Monte Carlo simulations were run using 1,000 participants in each trial.

Results

Figure 2 shows the three different simulations run based on the different estimated within-person standard deviation for the PRF Desirability scale. The y-axis label describes the possible bias that insufficient effort responding has on the correlation rather than the true correlation while the x-axis depicts the various proportions of valid respondents. The simulation data using Holtzman and Donnellan's (2017) random respondent simulator showed that the relationship approached zero after removing invalid respondents. The strongest correlation (r = -.34 for .15 within-person SD; r = -.30 for 0.20 within-person SD; r = -.27 for 0.25 within-person SD) occurred when 50% of the participants were simulated to be invalid. This means that the bias was most extreme when the study involved 50% invalid participants.

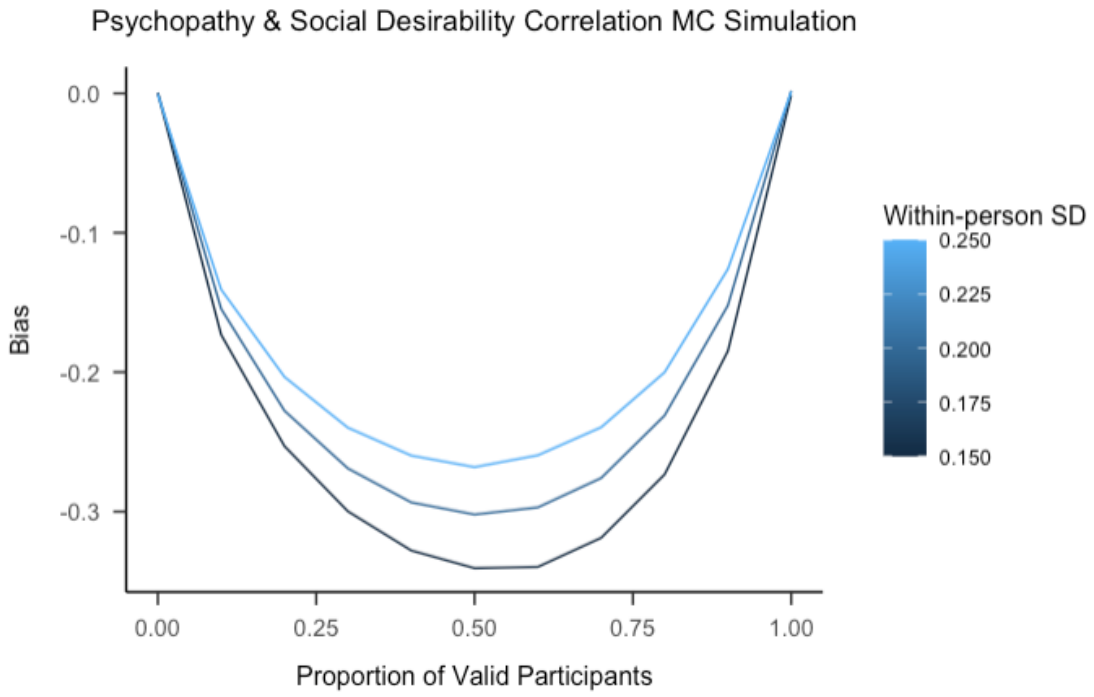

Figure 2. This graph shows the Monte Carlo simulation data based on the three different estimated within-person standard deviation for the PRF Desirability scale.

Discussion for Study 1

The simulated data showed that even 10% of the data being invalid could misrepresent the relationship between two variables because the correlation went from between -.18 at 10% invalid respondents with 0.15 as the within-person standard deviation, -.15 with 0.20 as the within-person standard deviation, and -.13 with 0.25 as the within-person standard deviation to nearly zero. This is similar to what Huang and colleagues (2015) had concluded from their research; even 5% of insufficient effort responding in the data can bias the relationship between two variables. When 100% of the participants were deemed valid, the correlation approached zero. This tells us that insufficient effort responding could potentially be biasing the relationship between psychopathy and social desirability. The results of this simulation provide support for further investigation into the relationship between psychopathy and social desirability under

insufficient effort responding. Had the correlation not decreased to zero as the proportion of valid participants increased, there may not be a need to conduct a further analysis. However, insufficient effort responding does seem to be biasing this relationship; therefore, a real-life correlational study was warranted.

CHAPTER 3

STUDY 2

Based on the simulation, the data showed that the increase in valid participants attenuated the most extreme correlation (which was evident when there were only 50% of respondents who provided valid [non-IER] data vs. 60%) between self-reported psychopathy and social desirability. The relationship became nearly non-existent when the data consisted of 100% valid respondents. Thus, this warranted investigating this relationship among human participants. Because of financial constraints, I was only able to collect data from 300 participants rather than the 1,000 participants used in the simulation. Participants took both the Self-report Psychopathy scale III (Paulhus et al., 2016; Williams et al., 2007) and PRF Desirability scale (Jackson, 1984), and correlations between the two scales were examined with and without insufficient effort responding data. In order to conduct the correlation without invalid data, various techniques were used to detect and eliminate insufficient effort responding. This study is also pre-registered on the Open Science Foundation website (https://osf.io/bkg87), and the pre-registered plan was followed.

Participants

Three hundred participants were recruited from a national adult community sample through Amazon Mechanical Turk, an online survey system. Only participants 18 years of age or older were included in the study. Participants were compensated $1.45 for completing the study. Three participants were excluded due to their request that their data not be used, and one participant was excluded due to missing data. The final sample consisted of 296 adults (40.5% female, n = 120). After insufficient effort responding analysis, 218 participants were deemed as invalid respondents, 78 participants were considered valid. Participants were between the ages of 19 and 71 years old (M = 36.61, SD = 11.06). The majority of participants self-identified as

White (82.1%, n = 243), 10 identified as Black (3.4%), 35 identified as Asian (11.8%), five participants identified as two or more races (1.7%), and three identified as American Indian or Alaskan Native (1%). Sixty-seven of the participants identified as Hispanic or Latino (22.6%).

Procedure

The Qualtrics link for the measures was available to participants through Amazon Mechanical Turk. Participants began by reading the informed consent form and provided their electronic consent by checking the option that they have read the consent form and agree to participate in the study. Participants were masked from the true hypothesis and purpose of the study, so a debriefing form was provided at the end of the survey explaining the purpose and hypothesis of the study, and participants had the option to exclude their data from the study after debriefing. Because this was an anonymous survey, participants did not provide their names on the consent or debriefing form. They completed the Self-report Psychopathy scale III (Paulhus et al., 2016; Williams et al., 2007) as well as the Desirability subscale from the Personality Research Form (Jackson, 1984), and the measures were counterbalanced.

Materials

Self-report Psychopathy Scale III. The Self-report Psychopathy scale III (SRP-III) consists of 64 items measured on a five-point Likert scale (1 = *disagree strongly* to 5 = *agree strongly*) (Paulhus et al., 2016; Williams et al., 2007). The SRP-III is also divided into four subscales to reflect the suggested four-factor structure of the PCL-R (Gordts et al., 2017; Hare, 2003). The subscales are Interpersonal Manipulation (reflecting the interpersonal factor that draws from characteristics like manipulation and pathological lying), Callous Affect (reflects the affective factor which relates to characteristics such as lack of remorse and empathy), Erratic Lifestyle (reflects the lifestyle factor that describes impulsivity and recklessness), and Criminal Tendencies

(reflects the antisocial factor which includes inclination toward antisocial behavior and criminality) (Gordts et al., 2017).

The SRP-III has converged with other well-validated self-report psychopathy measures such as the Psychopathy Personality Inventory Revised (Lilienfeld & Widows, 2005) and the Levenson Self-report of Psychopathy (Levenson et al., 1995) and negatively correlated with agreeableness on a self-report measure of the Five Factor Model of personality (Seibert et al., 2011). The SRP-III shows good construct validity through correlations with other psychopathy-related constructs such as antisocial behavior and low levels of empathy (Mahmut et al., 2011). In previous research, the Cronbach's alpha for the four subscales ranged from .69 to .82, and Cronbach's alpha for the overall scale was .90 (Seibert et al., 2011). In my study, Cronbach's alpha within the valid respondents was .89; within the insufficient effort respondents, the alpha was .96; ignoring insufficient effort responding status, the alpha was .95.

Personality Research Form Desirability Scale. The Desirability subscale comes from the Personality Research Form [Form E; Jackson (1984)]. The scale items consist of statements that are either deemed socially desirable or socially undesirable, and participants will choose to either endorse the statement about themselves by choosing *True* or not endorse the statement by choosing *False*. The scale is designed to measure one's tendency to respond in a socially desirable manner. Although cutoff scores have not been established, relatively high scores indicate possible impression management and conscious misrepresentation, and low scores may indicate a tendency to malinger or possibly low self-regard (MacNeil & Holden, 2006). The PRF Desirability scale has shown positive association with other social desirability scales (Holden & Fekken, 1989). Specifically, the Edwards (1957) Social Desirability scale and the PRF Desirability scale shared a similar dimension of social desirability (Holden & Fekken, 1989). Previous research shows that the Cronbach's alpha for the PRF Desirability scale ranged from .61

to .73 (Helmes & Holden, 2003; Holden & Fekken, 1989; Holden & Passey, 2010). In Study 2, Cronbach's alpha within the valid respondents was -.33; within the insufficient effort respondents, the alpha was .72; ignoring insufficient effort responding status, the alpha was .66.

Data Analysis

Using suggested methods by Curran (2016) to detect insufficient effort responding, IER detection began with the observation of response time. Response time was recorded for each participant, and pilot testing was done to estimate the average time it takes to complete both surveys. To account for possible outliers due to the nature of online surveys, the median response time was used as the reference point. Participants finishing faster than half of the median time (median = 479 seconds; IER criteria < 240) it takes to complete the survey were eliminated. This is because both surveys require individuals to carefully read the questions and then respond. Reading and responding in a thoughtful manner require time; therefore, individuals that do not seem to be taking the time to read the questions can be assumed to have provided invalid data. Sixty-five participants had response times less than 240 seconds.

Further detection was done through long string analysis (Huang et al., 2012; Meade & Craig, 2012). Long string analysis, also known as response pattern indices, is the analysis of a string of response in detection for careless or insufficient effort responding. This technique looks for the longest string of the same response from a participant, and it is based on the assumption that individuals who respond carelessly will choose the same option to every question (Huang et al., 2012; Meade & Craig, 2012). There are different ways in employing this technique. One way is to only detect for the longest single response option, and the other is to create a cutoff for each response option (Huang et al., 2012; Johnson, 2005; Meade & Craig, 2012). For this analysis, I detected only the longest single response option. Participants that chose the same option 5 times or more in a row were isolated. Sixty-seven SRP long string responses were detected, and 105

were detected for the PRF Desirability scale. If either analysis found respondents that had selected the same answer five times or more in a row, then participants were categorized as careless respondents. The long-string detection could have also overlapped. This means a participant with long-string responses on the SRP may also have long-string responses on the PRF Desirability scale.

Mahalanobis distance analysis was used to assess for outliers in the data. Mahalanobis distance is an analysis that examines for outliers in a multivariate space (De Maesschalck et al., 2000). In the conventional analysis of outliers, they are calculated by measuring the distance that a value is away from the middle of a distribution (Curran, 2016). However, the Mahalanobis distance analysis employs a two-dimensional technique that examines the combination of two or more variables and its distance away from the center of the multivariate distribution. Additionally, it also takes into account the covariance of the variables (Curran, 2016). This technique correlates well with other techniques used to detect insufficient effort responding (Maniaci & Rogge, 2014; Meade & Craig, 2012). Participants with greater than four Mahalanobis distance squared was isolated. Twenty-eight participants met the criteria and thus were categorized as insufficient effort responding.

Lastly, an odd-even consistency correlation was examined to measure individual reliability. Although both Cronbach's alpha and odd-even consistency are both used to measure reliability, odd-even consistency is being used to assess individual reliability here rather than scale reliability. Odd-even consistency is calculated by dividing participants' responses into their respective subscales with one set of odd numbered items and one set of even numbered items (Curran, 2016). The split odd or even items are then averaged for an approximate individual score on that subset of items. Then, these subscale subsets are divided into their respective odd or even vectors, and these two vectors are then correlated. The correlation coefficient can be corrected

using the Spearman-Brown formula if length of the vectors is too short (Curran, 2016). No participants were excluded due to a weak odd-even consistency correlation. In sum, insufficient effort responding detection were determined by response time, long-string analysis, Mahalanobis distance, and odd-even consistency correlation. Because all of these techniques had a firm reference point, there was no flexibility in the removal of insufficient effort respondents. It is important to note that some of the insufficient effort responding criteria did overlap. For example, participants who were isolated due to long string analysis may also have a response time less than 240 seconds. The final sample consisted of 218 invalid respondents and 78 valid respondents.

First, I performed a correlation analysis with all data included. Then, after eliminating the careless respondents from the data using the techniques listed above, I conducted another correlation analysis without the careless respondents. Because this is in part a replication from the Kowalski et al. (2018) study, this correlation should be replicated and should be significant in a negative direction. If the correlation with all data included has a correlation closer to zero than the correlation without insufficient effort responding, then data analysis would have stopped there because this would imply that the negative relationship between psychopathy and social desirability previously found was not biased due to insufficient effort responding. However, because correlation without insufficient effort responding is closer to zero than the correlation with all data and is non-significant, I conducted an equivalence test to determine whether the effect (in the valid-only data) is statistically equivalent to zero (Lakens, 2017).

Equivalence testing can either be used to examine whether a relationship that is extreme enough to become significant can be rejected or whether there really is an absence of a relationship between variables. This is because with a large enough sample size, a weak but significant relationship may be detected (Lakens et al., 2018). Additionally, small sample sizes may not always yield significant relationships due to low power. However, equivalence testing

can help with determining whether there is really a relationship worth examining (Lakens, 2017). Equivalence testing is typically done using two one-sided tests (also known as the TOST procedure) with a one-sided test for the lower bound and a one-sided test for the upper bound. These boundary values are determined based on the expected difference between the populations being examined. If the observed data shows a difference outside of this expected difference, then it is possible that there is an effect extreme enough worth examining (Lakens, 2017; Lakens et al., 2018).

The null hypothesis for the TOST procedure is the presence of a true relationship that falls outside of the specified boundary values. The alternative hypothesis is that when comparing the observed data to the boundary values specified for the TOST procedure, the observed relationship falls within those boundaries and thereby shows an absence of a meaningful relationship worth examining (Lakens, 2017). A TOST p-value below .05 signifies that the effect is essentially equivalent to zero; however, both one-sided tests need to be significant to draw a conclusion of statistical equivalence (Lakens et al., 2018).

Results

As expected, in the all-inclusive (IER and non-IER) data, the association between psychopathy (M = 2.74, SD = .62) and social desirability (M = .64, SD = .18) was significantly negatively correlated (r = -.49, p < .001). The correlation between psychopathy (M = 2.63, SD = .65) and social desirability (M = .67, SD = .20) within IER only data was r = -.48, p < .001. After insufficient effort responding was eliminated from the dataset, the association (r = -.22, p = .054) between psychopathy (M = 3.03, SD = .39) and social desirability (M = .57, SD = .10) became non-significant and was attenuated by .20. Because the correlation without insufficient effort responding was attenuated, the TOST procedure was performed to test for statistical equivalence (Lakens et al., 2018). The TOST remained non-significant (lower bound p = .198, upper bound p

< .001). Because both one-sided tests need to be significant for the TOST to be considered significant, it can be concluded that the observed relationship between psychopathy and social desirability is statistically different from zero according to the TOST procedure. However, the combination of the non-significant hypothesis test for valid data and non-significant TOST yields inconclusive results. Figure 3 shows the correlation between psychopathy and social desirability scores in both insufficient effort responding and non-insufficient effort responding data followed by the TOST results.

Figure 3. This figure shows the correlation between average psychopathy and social desirability scores. The red dots represent insufficient effort responding data, and the blue dots represent non-insufficient effort responding data.

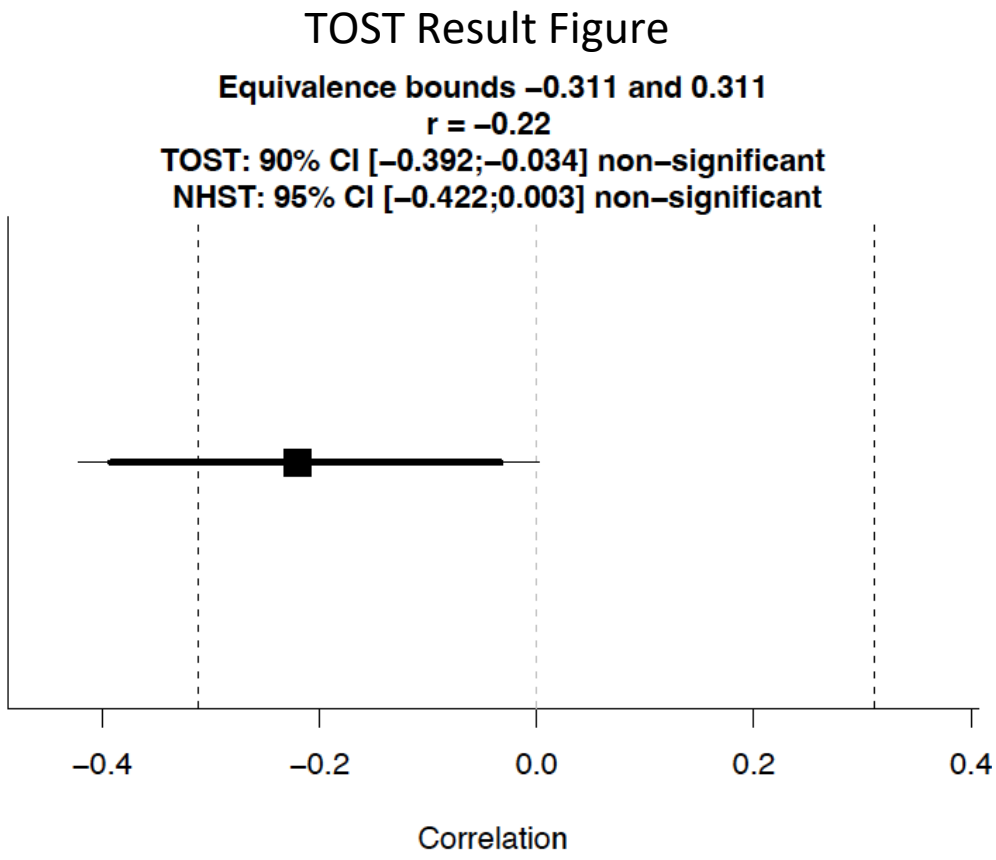

Figure 4. This is the TOST results figure. The darker dotted lines on either side represent the lower and upper boundaries. The black line with the square represents the correlation between psychopathy and social desirability and its confidence interval. The dark flat line partly being outside of the lower boundary signifies a non-significant equivalence test. The lower and upper bound of the hypothesis significance testing confidence interval would both need to be within the boundaries of the dark dotted lines for the TOST to be considered significant.

Discussion for Study 2

I predicted that the relationship between psychopathy and social desirability will approach zero after eliminating insufficient effort responding data and that the TOST procedure will provide further confirmation that there is no relationship between psychopathy and social desirability. This study did find that the relationship between psychopathy and social desirability became non-significant after isolating the insufficient effort responding data. However, the TOST

procedure produced a contradicting conclusion. The combination of the hypothesis test and the TOST yielded inconclusive results.

The careless package (https://github.com/ryentes/careless/) was used to detect insufficient effort responding data through odd-even consistency, long string analysis, and Mahalanobis distance. In order to analyze long string responses, the items within the two measures remained non-randomized. However, this may not have been an ideal presentation of the PRF Desirability scale. The Desirability scale was scored in a way that if the questions were presented in order and participants responded *true* on every other question, it would have been a perfect high score. With a number of perfect high scores within the insufficient effort responding data, this could have contributed to the biased correlation. The long string analysis of the non-randomized presentation of the PRF Desirability scale could have also disqualified some valid participants as invalid, and this calls into question the validity of the *valid* correlation found between the two variables. For example, if participants answered *true* five times in a row, they would have been deemed as invalid participants even though these responses could have been true and honest. Randomization of the PRF Desirability scale is highly recommended for future use.

Though the valid-only data presented a correlation magnitude nearly identical to that of Kowalski et al. (2018) (N = 244, r = -.20, p <. 01), the correlation within Study 2 was non-significant (N = 78, r = -.22, p = .054). The result of the null hypothesis testing contradicted the result of the equivalence testing. According to Lakens et al. (2018), this combination indicates that more investigation into the relationship is needed. A larger sample size may deliver the statistical power needed to detect a significant relationship. Because the equivalence bounds are based on the smallest effect size of interest, as the smallest effect size of interest decreases, the equivalence bound also becomes narrower (Lakens, 2017; Lakens et al., 2018). Similar to

hypothesis testing, a larger sample size would help narrow the confidence interval to provide detection power (Lakens et al., 2018).

Lakens (2017) notes that a non-significant relationship in the traditional null hypothesis testing does not necessarily indicate there is no effect; thus, accepting the null would be problematic. This is why equivalence testing is important. Equivalence testing can help determine whether an effect worth examining (statistically different from zero) should be rejected (if statistically equivalent to zero) or whether a non-significant effect (not statistically different from zero) is worth looking into. Because the null hypothesis in equivalence testing is the presence of a true effect, not rejecting the null (a non-significant TOST) in this case actually contradicts the conclusion that the correlation is not statistically different from zero (the absence of a relationship). Hence, additional research into this relationship is necessary.

CHAPTER 4

GENERAL DISCUSSION

Because psychopathic characteristics include manipulation, egocentricity, and deception, I hypothesized that psychopathy would be positively associated with socially desirable responding when completing self-report surveys (Miller et al., 2011; Neumann & Hare, 2008). The consensus in the literature is that there is a significant negative relationship between social desirability and psychopathy, and this also implies that the psychopathy scores found within the literature are a reflection of their true psychopathy scores (Kowalski et al., 2018; Ray et al., 2013; Verschuere et al., 2014). Still, some studies did not find that those higher on the psychopathy scale were less likely to engage in socially desirable responding (Book et al., 2006; Fisher et al., 2018).

A meta-analysis that examined the relationship between psychopathy and socially desirable responding did find positive (though non-significant) relationships between certain components of psychopathy and social desirability (Ray et al., 2013). The overall conclusion from the meta-analysis, however, was that the relationship between psychopathy and social desirability was either negative or negligible. I suspected that insufficient effort responding could artificially push the correlation between psychopathy and social desirability toward the negative pole, and I hypothesized that this relationship may be negligible after eliminating careless responding.

In the simulation, after increasing the proportion of valid participants, the relationship between psychopathy and social desirability was attenuated and eventually became negligible after the proportion of valid participants reached 100%. In Study 2, although the correlation between psychopathy and social desirability went in the expected negative direction, it may not have been for the reasons I expected. Figure 3 shows the insufficient effort responding data

points along with the non-insufficient effort responding data points. The red dots, representing insufficient effort responding data, are scattered near the top left of the graph, and the blue dots, representing non-insufficient effort responding data, are scattered near the center of the graph. This is the opposite of what I expected where the data points would scatter as depicted by Figure 2. One way to explain the red dots would be that participants answered in a patterned way (i.e., selecting only 1 & 2). Although insufficient effort responding can produce non-random patterns, it is often thought that the responses would center the midpoint of the scale after averaging the scores (Huang et al., 2012). This was a particular pattern that was not expected. The blue dots also deviated from expectation in that it is centered around the midpoint of the scales. This could simply be that this sample of participants were moderate in psychopathy and social desirability.

One limitation in this study that also contributes to the inconclusiveness of the results is that the Cronbach's alpha for the valid PRF Desirability data was negative. The unreliable measure makes it difficult to draw any definitive conclusion. It is possible that a meaningful relationship may be observed through different types of measurement. Given the studies that found a positive relationship, it is still possible that when given good enough incentives, people high in psychopathy will try to present themselves as socially desirable in order to obtain the incentives (Book et al., 2006; Fisher et al., 2018; Ray et al., 2013). However, in terms of self-report measures of psychopathy and social desirability without incentives, there may need additional exploration of this relationship.

Conclusion

Overall, the results of the studies yielded inconclusive data that careless responding may be biasing the relationship between psychopathy and social desirability. However, because there are studies that found a positive relationship between faking behaviors and psychopathy, this relationship may be dependent on the method of measurement and setting (Book et al., 2006;

Fisher et al., 2018; Ray et al., 2013). In conclusion, further examination is needed into the relationship between self-report psychopathy and socially desirable responding.

REFERENCES

Beach, D. A. (1989). Identifying the random responder. *The Journal of Psychology*, *123*(1), 101–103. https://doi.org/10.1080/00223980.1989.10542966

Beussink, C. N., Hackney, A. A., & Vitacco, M. J. (2017). The effects of perspective taking on empathy-related responses for college students higher in callous traits. *Personality and Individual Differences*, *119*, 86–91. https://doi.org/10.1016/j.paid.2017.07.005

Book, A. S., Holden, R. R., Starzyk, K. B., Wasylkiw, L., & Edwards, M. J. (2006). Psychopathic traits and experimentally induced deception in self-report assessment. *Personality and Individual Differences*, *41*(4), 601–608. https://doi.org/10.1016/j.paid.2006.02.011

Cleckley, H. (1941). *The mask of sanity*. The CV Mosby Company. https://doi.org/10.1080/00325481.1951.11694097

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Curran, P., & Hauser, D. (2015). Understanding responses to check items: A verbal protocol analysis. *Philadelphia, PA: Paper Presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology*.

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Dryden Press.

Edwards, A. L. (1958). The social desirability variable in personality assessment and research.

Academic Medicine, *33*(8), 610–611. https://journals.lww.com/academicmedicine/Citation/1958/08000/The_Social_Desirability_Variable_in_Personality.24.aspx

Fisher, P. A., Robie, C., Christiansen, N. D., & Komar, S. (2018). The impact of psychopathy and warnings on faking behavior: A multisaturation perspective. *Personality and Individual Differences*, *127*, 39–43. https://doi.org/10.1016/j.paid.2018.01.033

Gordts, S., Uzieblo, K., Neumann, C., Van den Bussche, E., & Rossi, G. (2017). Validity of the self-report psychopathy scales (SRP-III full and short versions) in a community sample. *Assessment*, *24*(3), 308–325. https://doi.org/10.1177/1073191115606205

Groth, J., & Kleka, P. (2018). Patterns of intentional faking in questionnaire-based study of psychopathy. *Current Issues in Personality Psychology*, *6*(4), 305–317. https://doi.org/10.5114/cipp.2018.80199

Hare, R. D. (1991). *The psychopathy checklist–revised (PCL-r)*. Multi-Health Systems.

Hare, R. D. (2003). *The psychopathy checklist–revised (PCL-r)* (2nd ed.). Multi-Health Systems.

Hare, R. D., & Neumann, C. S. (2005). Structural models of psychopathy. *Current Psychiatry Reports*, *7*(1), 57–64. https://doi.org/10.1007/s11920-005-0026-3

Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Helmes, E., & Holden, R. R. (2003). The construct of social desirability: One or two dimensions? *Personality and Individual Differences*, *34*(6), 1015–1023. https://doi.org/10.1016/S0191-8869(02)00086-7

Holden, R. R., & Fekken, G. C. (1989). Three common social desirability scales: Friends, acquaintances, or strangers? *Journal of Research in Personality*, *23*(2), 180–191.

Holden, R. R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences*, *49*(5), 446–450. https://doi.org/10.1016/j.paid.2010.04.015

Holtzman, N. S., & Donnellan, M. B. (2017). A simulator of the degree to which random responding leads to biases in the correlations between two individual differences. *Personality and Individual Differences*, *114*, 187–192. https://doi.org/10.1016/j.paid.2017.04.013

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828. https://doi.org/10.1037/a0038510

Jackson, D. N. (1984). *Personality research form manual (3rd ed.)*. Sigma Assessment Systems.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and amazon's mechanical turk. *Journal of Advertising*, *46*(1), 141–155. https://doi.org/10.1080/00913367.2016.1269304

Kowalski, C. M., Rogoza, R., Vernon, P. A., & Schermer, J. A. (2018). The dark triad and the self-presentation variables of socially desirable responding and self-monitoring. *Personality and Individual Differences*, *120*, 234–237. https://doi.org/10.1016/j.paid.2017.09.007

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177%2F1948550617697177

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177%2F2515245918770963

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*(1), 151. https://psycnet.apa.org/doi/10.1037/0022-3514.68.1.151

Lilienfeld, S. O., & Arkowitz, H. (2007). What "psychopath" means. *Scientific American Mind, 18*(6), 80–81. http://faculty.fortlewis.edu/burke_b/forensic/Class%20Readings/Psychopathy.pdf

Lilienfeld, S. O., & Widows, M. (2005). *Professional manual for the psychopathic personality inventory-revised (PPI-r)*. Psychological Assessment Resources.

Lynam, D. R., Gaughan, E. T., Miller, J. D., Miller, D. J., Mullins-Sweatt, S., & Widiger, T. A. (2011). Assessing the basic traits associated with psychopathy: Development and validation of the elemental psychopathy assessment. *Psychological Assessment*, *23*(1), 108.

MacNeil, B. M., & Holden, R. R. (2006). Psychopathy and the detection of faking on self-report inventories of personality. *Personality and Individual Differences*, *41*(4), 641–651. https://doi.org/10.1016/j.paid.2006.03.004

Mahmut, M. K., Menictas, C., Stevenson, R. J., & Homewood, J. (2011). Validating the factor structure of the self-report psychopathy scale in a community sample. *Psychological Assessment*, *23*(3), 670. https://doi.org/10.1037/a0023090

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437. https://doi.org/10.1037/a0028085

Meffert, H., Gazzola, V., Den Boer, J. A., Bartels, A. A., & Keysers, C. (2013). Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. *Brain*, *136*(8), 2550–2562. https://doi.org/10.1093/brain/awt190

Miller, J. D., Gaughan, E. T., Maples, J., Gentile, B., Lynam, D. R., & Widiger, T. A. (2011). Examining the construct validity of the elemental psychopathy assessment. *Assessment*, *18*(1), 106–114. https://doi.org/10.1177/1073191110393139

Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PloS One*, *11*(6), e0157732. https://doi.org/10.1371/journal.pone.0157732

Neumann, C. S., & Hare, R. D. (2008). Psychopathic traits in a large community sample: Links to violence, alcohol use, and intelligence. *Journal of Consulting and Clinical Psychology*, *76*(5), 893. https://doi.org/10.1037/0022-006X.76.5.893

Neumann, C. S., Vitacco, M. J., Hare, R. D., & Wupperman, P. (2005). Reconstruing the "reconstruction" of psychopathy: A comment on cooke, michie, hart, and clark. *Journal of Personality Disorders*, *19*(6), 624–640. https://doi.org/10.1521/pedi.2005.19.6.624

Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2016). *Manual for the self-report psychopathy scale*. Multi-Health Systems.

Pérez, B., Herrero, J., Velasco, J., & Rodriguez-Díaz, F. J. (2015). A contrastive analysis of the

factorial structure of the PCL-r: Which model fits best the data? *European Journal of Psychology Applied to Legal Context*, *7*(1), 13–30. https://doi.org/10.1016/j.ejpal.2014.10.001

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, *78*(3), 582. https://doi.org/10.1037/0022-3514.78.3.582

Ray, J. V., Hall, J., Rivera-Hudson, N., Poythress, N. G., Lilienfeld, S. O., & Morano, M. (2013). The relation between self-reported psychopathic traits and distorted response styles: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, *4*(1), 1. https://doi.org/10.1037/a0026482

Seibert, L. A., Miller, J. D., Few, L. R., Zeichner, A., & Lynam, D. R. (2011). An examination of the structure of self-report psychopathy measures and their relations with general traits and externalizing behaviors. *Personality Disorders: Theory, Research, and Treatment*, *2*(3), 193. https://doi.org/10.1037/a0019232

Verschuere, B., Uzieblo, K., De Schryver, M., Douma, H., Onraedt, T., & Crombez, G. (2014). The inverse relation between psychopathy and faking good: Not response bias, but true variance in psychopathic personality. *The Journal of Forensic Psychiatry & Psychology*, *25*(6), 705–713. https://doi.org/10.1080/14789949.2014.952767

Vitacco, M. J., Rogers, R., Neumann, C. S., Harrison, K. S., & Vincent, G. (2005). A comparison of factor models on the PCL-r with mentally disordered offenders: The development of a four-factor model. *Criminal Justice and Behavior*, *32*(5), 526–545. https://doi.org/10.1177/0093854805278414

Watts, A. L., Lilienfeld, S. O., Edens, J. F., Douglas, K. S., Skeem, J. L., Verschuere, B., &

LoPilato, A. C. (2016). Does response distortion statistically affect the relations between self-report psychopathy measures and external criteria? *Psychological Assessment, 28*(3), 294. https://doi.org/10.1037/pas0000168

Williams, K. M., Paulhus, D. L., & Hare, R. D. (2007). Capturing the four-factor structure of psychopathy in college students via self-report. *Journal of Personality Assessment, 88*(2), 205–219.

Wilson, L., Miller, J. D., Zeichner, A., Lynam, D. R., & Widiger, T. A. (2011). An examination of the validity of the elemental psychopathy assessment: Relations with other psychopathy measures, aggression, and externalizing behaviors. *Journal of Psychopathology and Behavioral Assessment*, *33*(3), 315–322.

Woodworth, M., & Porter, S. (2002). In cold blood: Characteristics of criminal homicides as a function of psychopathy. *Journal of Abnormal Psychology*, *111*(3), 436. https://doi.org/10.1037//0021-843x.111.3.436