

6-2014

IMPROVING DATA QUALITY WITH FOUR SHORT SENTENCES: HOW AN HONOR CODE CAN MAKE THE DIFFERENCE DURING DATA COLLECTION

Alexander Scott McKay
California State University - San Bernardino

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/etd>



Part of the [Industrial and Organizational Psychology Commons](#), [Personality and Social Contexts Commons](#), [Quantitative Psychology Commons](#), and the [Social Psychology Commons](#)

Recommended Citation

McKay, Alexander Scott, "IMPROVING DATA QUALITY WITH FOUR SHORT SENTENCES: HOW AN HONOR CODE CAN MAKE THE DIFFERENCE DURING DATA COLLECTION" (2014). *Electronic Theses, Projects, and Dissertations*. 78.

<https://scholarworks.lib.csusb.edu/etd/78>

This Thesis is brought to you for free and open access by the Office of Graduate Studies at CSUSB ScholarWorks. It has been accepted for inclusion in Electronic Theses, Projects, and Dissertations by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

IMPROVING DATA QUALITY WITH FOUR SHORT SENTENCES:
HOW AN HONOR CODE CAN MAKE THE DIFFERENCE
DURING DATA COLLECTION

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Psychology:
General Experimental

by
Alexander Scott McKay
June 2014

IMPROVING DATA QUALITY WITH FOUR SHORT SENTENCES:
HOW AN HONOR CODE CAN MAKE THE DIFFERENCE
DURING DATA COLLECTION

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

by

Alexander Scott McKay

June 2014

Approved by:

Dr. Donna M. Garcia, Committee Chair, Psychology

Dr. Kenneth S. Shultz, Committee Member

Dr. John Clapper, Committee Member

Dr. James C. Kaufman, Committee Member, University of Connecticut

© 2014 Alexander Scott McKay

ABSTRACT

Careless responders have a large impact on a study by causing issues such as Type II errors (failing to reject a false null hypothesis), which then waste researchers' time and money. Research on careless responding has focused primarily on detecting and removing careless responders rather than on reducing careless responding before data collection begins. The purpose of the present study was to test the use of honor codes with or without the presentation of a picture of watchful eyes to increase self-awareness thereby reducing careless responding. Participants ($N = 305$) were randomly assigned to one of five honor code conditions (control condition, read-only condition, type condition, read-and-eyes condition, or type-and-eyes condition) and then completed a number of personality measures. Participant's responses were screened for careless responding. I found that when participant's only read an honor code without a picture of watchful eyes on the screen, they were significantly more likely to engage in careless responding than were people in the control condition. There was no significant difference in careless responding in the control condition compared to the other three conditions (type condition, read-and-eyes condition, or type-and-eyes condition). This finding indicates that participants being presented only with an honor code, and no other cues of moral behavior, might lead to psychological reactance (i.e., participants perceiving their freedom or control was threatened). This psychological reactance was likely due to participants feeling coerced to respond honestly by the honor code rather than

participants responding honestly on their own accord. Additionally, there were no differences on participant's responses across conditions on a number of personality measures that may be sensitive to increased self-awareness. A number of exploratory analyses were also conducted to further examine psychological reactance. The present study provides answers and possible directions to a number of useful questions that could improve data quality and reduce potential negative effects of honor codes under certain circumstances.

ACKNOWLEDGMENTS

I would like to thank Jackie McConnaughy and my family. Their support has helped me accomplish my goals in both school and life.

I would also like to thank Bob Cupp, Dr. James Kaufman, and Dr. Donna Garcia. I would like to thank Bob Cupp for being one of the first professors to challenge me as a student and inspire me to pursue graduate school and a career in research. If it were not for his passion, I would have likely never considered pursuing research. I would also like to thank Dr. James Kaufman and Dr. Donna Garcia. The two of them invited me into their research labs and gave me so many opportunities that I could barely keep up with all of them! Their dedication to help me be successful in school and in my future endeavors will always be remembered and appreciated. I cannot express in words my appreciation for their generosity.

Last, I would like to thank Michelle Fabros for helpful comments while working on this project, Heather Roessler for her assistance with data collection, Courtney Shields for her helpful comments on an early draft, and my thesis committee for their helpful comments and advice.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	v
CHAPTER ONE: INTRODUCTION	
Introduction	1
What is Careless Responding?	4
Assessment of Careless Responding	8
Explicit Measures for Detecting Careless Responding	9
Post Hoc Data Screening Methods	11
Summary of Screening Methods	13
Base Rates of Careless Responding	14
Previous Attempts to Reduce Careless Responding	17
Honor Codes and Watchful Eyes: Their Effects on Behavior	20
Honor Codes	21
Watchful Eyes	22
Honor Codes and Watchful Eyes	22
The Present Study	23
Hypothesis 1	25
Hypothesis 1a	25
Hypothesis 1b	25
Hypothesis 1c	25
Hypothesis 2	25

Hypothesis 2a	26
Hypothesis 2b	26
Research Question #1	26
Research Question #2	26
Research Question #3	27
Research Question #4	27
Research Question #5	28
Research Question #6	29
 CHAPTER TWO: METHOD	
Participants and Recruitment	30
Materials	31
Measures to Reduce Careless Responding	32
Measures to Detect Careless Responding	32
Personality Measures	33
HEXACO-100	34
Machiavellianism (MACH-IV)	34
Domain-Specific Risk-Taking (DOSPERT)	35
Right Wing Authoritarianism (RWA)	36
Social Dominance Orientation (SDO)	36
Self-Esteem.....	37
Balanced Inventory of Desirable Responding (BIDR)...	37
Self-Reported Psychopathy-III (SRP-III).....	38

Procedure	39
 CHAPTER THREE: RESULTS	
Sample	42
Handling of Duplicate Cases and Missing Data	42
Test of Hypotheses and Research Questions	44
Reduction in Careless Responding among Conditions	44
Hypothesis 1a, 1b, and 1c and Research Questions #1 and #2	44
Hypothesis 2	48
Research Question #3	49
The Effects of an Honor Code and Watchful Eyes on Survey Responses	51
Research Question #4	51
Correlations among Careless Responding Indices, Participant Engagement, and Personality Measures	53
Correlations among Careless Responding Indices and Participant Engagement	53
Correlations among Careless Responding Indices and Personality Measures	55
Correlations among Participant Engagement and Personality Measures	56
Method of Completing Study and Technological Issues	58
Post Hoc Analyses to Research Question #2 and Research Question #5	61
 CHAPTER FOUR: DISCUSSION	
Introduction	63

Conclusions to Hypotheses and Research Questions and Directions for Future Research	63
Limitations of Study Design and Procedures	76
Conclusion	79
APPENDIX A: INFORMED CONSENT FORM	81
APPENDIX B: HONOR CODE AND WATCHFUL EYES	83
APPENDIX C: CARELESS RESPONDING ITEMS	85
APPENDIX D: PARTICIPANT ENGAGEMENT MEASURES	87
APPENDIX E: HEXACO-100	89
APPENDIX F: MACHIAVELLIANISM (MACH-IV)	94
APPENDIX G: DOMAIN-SPECIFIC RISK-TAKING (DOSPERT)	96
APPENDIX H: RIGHT WING AUTHORITARIANISM (RWA)	99
APPENDIX I: SOCIAL DOMINANCE ORIENTATION (SDO)	102
APPENDIX J: SELF-ESTEEM.....	104
APPENDIX K: BALANCED INVENTORY OF DESIRABLE RESPONDING (BIDR)	106
APPENDIX L: SELF-REPORTED PSYCHOPATHY-III (SRP-III)	109
APPENDIX M: DEMOGRAPHICS	112
APPENDIX N: DEBRIEFING FORM	114
APPENDIX O: TABLES	116
APPENDIX P: FIGURES	124
REFERENCES	131

CHAPTER ONE

INTRODUCTION

Introduction

Researchers increasingly rely on online data collection strategies.

Although these strategies have many benefits, they also have their problems.

One of the largest problems is with the number of participants who do not provide thoughtful responses. Although there are new methods to identify these “careless responders” and sound arguments for why these cases should be removed from data sets, few strategies are available that minimize careless responding before data collection begins. In the proposed study, I sought to address the issue of careless responding, with a focus on how to increase thoughtful responding. I began with a brief discussion of online research methods. Then I defined what careless responding is and its effect on data quality. Next, I provided an overview of methods used to screen data sets for these responders, described the limitations of this approach, and presented rates of careless responding in research. I then discussed the research on the benefits of honor codes. Finally, I proposed a study to test the effects of honor codes and determine their effectiveness in reducing careless responding during online survey research.

The Internet has become a viable tool for data collection. Since Gosling, Vazire, Srivastava, and John (2004) addressed many of the concerns researchers originally had with the use of online samples, researchers have

flocked to the Internet with three primary goals. First, Internet research allows researchers to obtain hard-to-examine samples from around the world and within their own country (Buchanan & Smith, 1999; Gosling et al., 2004; Rentfrow, 2010). Second, Internet samples allow researchers to avoid a bias towards Western, Educated, Industrialized, Rich, and Democratic or “WEIRD” populations (Henrich, Heine, & Norenzayan, 2010). Third, by using the Internet for data collection, researchers can examine samples and populations other than typical college-aged participants, which for many reasons might not be representative of the general population (see Sears, 1986).

Researchers who are not interested in collecting these unique samples can still benefit from using the Internet to collect data. Gosling et al. (2004) argue that using the Internet can save researchers time because they do not need to enter data manually, and there is a reduced risk of coding errors during data entry. Empirical evidence suggests that these benefits can be gained while maintaining the data integrity associated with other data collection approaches. The switch to collecting data online has presented little difference in data quality relative to paper-and-pencil research involving surveys (Miller et al., 2002; Weigold, Weigold, & Russell, 2013) or behavioral tasks (Casler, Bickel, & Hackett, 2013) with an attentive sample. Consequently, samples from crowdsourcing websites (i.e., websites where people contract participants such as Amazon’s MTurk or Qualtrics’ Research Panels) are used regularly (and successfully) in experimental research (see Caruso, Vohs, Baxter, & Waytz,

2013; Wiltermuth & Gino, 2013), organizational research (Behrend, Sharek, Meade, & Wiebe, 2011; Stanton, 1998; Stanton & Rogelberg, 2001), and clinical research (Shapiro, Chandler, & Mueller, 2013). In addition to minimizing data entry time and errors, the benefits of using such samples can provide researchers with a more generalizable sample and increase external validity by moving the experiment out of the lab. Moreover, online studies are typically inexpensive, costing researchers pennies in the case of short surveys or slightly more for longer surveys (Buhrmester, Kwang, & Gosling, 2011; Mason & Suri, 2012). Buhrmester et al. (2011) demonstrated that a payment of 2, 10, or 50 cents did not result in a large change in Cronbach's alpha scores (less than a hundredth point change for most scales) for a number of personality measures (Big Five, Adult attachment).

Given their continued praises in research journals, online research appears to be without fault. These methods, however, have their limitations. For example, researchers lose the environmental control that ensures participants will complete the study in a distraction-free environment. The loss of this control, especially with online student samples, may cause some participants to devote less attention because they know that they will still get compensation. These participants' lack of effort could result in poor quality data, which can cause researchers problems when analyzing and publishing their work. These "careless responders" are one of the challenges presented with online research.

What is Careless Responding?

People use a number of different response patterns when completing a survey. First, and the most typical, a participant will complete a study by reading the instructions and survey items and answering truthfully. These participants are of little concern to researchers. Nichols, Greene, and Schmolck (1986) have identified two other response pattern biases that are of concern though. First, *content responsive faking* is when people respond in a manner that is influenced by the item's content and it is not an accurate reflection of the person completing the survey. This type of responding can be either "faking good" or "faking bad," depending on what the respondent is trying to accomplish. For example, a person on trial for a crime might "fake bad" in order to be deemed unfit for trial, whereas a person may "fake good" in order to appear more qualified for a job than they truly are. Tests such as the MMPI include measures to detect faking good and faking bad (Rogers, Sewell, Martin, & Vitacco, 2003). Measures of social desirability, such as Paulhus's (1991) Balanced Inventory of Desirable Responding (BIDR) self-deceptive enhancement subscale, have been used to detect participants attempting to fake good. Although these respondents are a problem for researchers, they were not the primary focus of this study.

The second type of response pattern bias identified by Nichols et al. (1986) is *content nonresponsivity*. Content nonresponsivity deals with a participant completing a study without reading instructions or survey items. These response types have been termed insufficient effort responding (Huang,

Curran, Keeney, Poposki, & DeShon, 2012), protocol invalidity (Johnson, 2005), careless responding (Curran, Kotrba, & Denison, 2010; Meade & Craig, 2012), and random responding (Baer, Ballenger, Berry, & Wetter, 1997; Beach, 1989; Berry et al., 1992; O'Dell, 1971). Meade and Craig (2012) argue that careless responding is a more apt term rather than random responding because people may not be responding in a random pattern. A participant may respond using a certain pattern (e.g., clicking the same box through the entire survey). The current study focused on reducing participant's engagement in content nonresponsivity.

There are five primary factors that might influence whether a person responds carelessly or not. Four of these factors have been addressed previously (see Meade & Craig, 2012) and one factor has not been addressed. First, participant interest might be an important factor for influencing the likelihood a participant will engage in careless responding. If participants are uninterested in the survey, they may be more likely to respond carelessly. This problem can be of greater concern in college samples. College students typically participate in research for reimbursement such as course credit. Other options, such as writing a paper, for extra credit (in lieu of research participation) are usually undesirable alternatives. If students feel research participation is the only reasonable option, they could respond malevolently, without interest, or without considering the implications of their careless responding.

Second, survey length can be a factor for careless responding. Some

surveys such as the MMPI-2 are over 500 items in length (Berry et al., 1992), and many online surveys contain a comparable number of items. Berry et al. (1992) found people self-reported being more likely to respond carelessly while answering questions towards the middle and end of the MMPI-2. Long surveys may cause individuals to feel fatigue, and they will rush through the survey in order to finish quickly. Barber, Barnes, and Carlson (2013) found that participants completing a study when tired were also more likely to report using less effort to complete the study.

Third, social contact may be important to reduce careless responding (Johnson, 2005; Meade & Craig, 2012). In online research, because participants complete the survey away from a lab, they gain increased anonymity. This total anonymity may cause people to complete a survey carelessly and view retribution (i.e., loss of extra credit points) as unlikely. Fourth, participants might be distracted while completing a survey outside of a laboratory. People might watch television, tend to a child, or text while completing an online survey. These and other types of distractions could reduce the attention needed to complete a study.

Fifth, people might complete the study on a tablet or a cellular phone, but the website is not tablet- or phone-compatible. People completing a study under these conditions might be inclined to either not complete a study or carelessly respond throughout the study. Data Collection websites (e.g., *Qualtrics*) have begun to develop mobile-compatible surveys with the use of responsive web

designs. Responsive web designs are web design setups that will adjust to the device that users are completing the study on (e.g., a tablet or cellular phone). Because companies have begun to address this potential issue, it might not be a concern for researchers within the next few years. For researchers independently developing their studies this issue might be a factor to consider when creating online surveys. Overall, any of the five factors individually or in combination could lead to greater careless responding in online research and hurt data quality.

Identifying careless responders is important for researchers to interpret their data accurately. One reason for identifying careless responders is the effect they can have on scale development. Schmitt and Stults (1985) found, using simulated data, that if 10 percent or more of participants are responding carelessly in a survey with both positive and negative worded items, a researcher might reject a one factor model using a principal components analysis. Differences in response patterns to the positively versus negatively worded items could falsely generate the appearance of a two-factor model. Woods (2006) also demonstrated the same effect using simulated data in a principal factor analysis. Huang et al. (2012), using nonsimulated data (i.e., personality surveys) collected from college students, found a similar occurrence with factor analysis. Maniaci and Rogge (2014; see also Oppenheimer, Meyvis, & Davidenko, 2009; Osborne & Blanchard, 2010) found that careless responders can also cause a large reduction in statistical power. This loss of power could lead a researcher to fail to reject a false null hypothesis and make a Type II error. The harmful effects that

careless responders have on data quality are of great concern for researchers given that a small percentage of participants responding carelessly can cause researchers to make false conclusions. These false conclusions might reflect issues with replication in psychological research in the lab and online.

Oppenheimer et al. (2009) noted that experiments with subtle manipulations in the wording of instructions or scenarios might result in a Type II error if participants are not completing a study attentively. To address this problem, methods to detect these cases have been developed.

Assessment of Careless Responding

Careless responding has seen a surge in research in the last decade (Barber et al., 2013; Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012; Oppenheimer et al., 2009; Osborne & Blanchard, 2010), but the interest in detecting careless responders is not new in psychological research. As noted by O'Dell (1971), researchers such as Raymond Cattell developed methods to detect response biases in personality research as early as the 1950s. These methods consisted of short scales to detect careless responding, with a focus on content responsive faking. Although these scale approaches were beneficial, with faster computers and data analysis programs, the methods have improved. Meade and Craig (2012) classify these methods into two categories. The first category is explicit in form and the second category involves post hoc data screening.

Explicit Measures for Detecting Careless Responding

Explicit measures to detect careless responding involve entire measures or single items at the beginning of or spread throughout a survey. Oppenheimer et al. (2009) developed instructional manipulation checks that flag participants who do not read a specific set of instructions. Oppenheimer and colleagues found this method to be effective in short experimental research in which participant responses are based on subtle changes in task wording (e.g., framing effects).

In survey research, other methods are typically used. These approaches include bogus items or instructed response items spread throughout a survey. Beach (1989) found that the inclusion of True/False items such as “I was born on February 30th” to detect careless responders was an effective method. Osborne and Blanchard (2010) found that using bogus items to detect careless responders was comparable to using the third parameter model of Item Response Theory (or the “guessing” parameter). Item Response Theory, however, is a more time intensive method than would be the use of bogus items. Chapman and Chapman’s (1983) unpublished Infrequency Scale uses the same principle. If a person fails two or more bogus items on the scale, his or her survey is deemed invalid and removed. Tests such as the MMPI and MMPI-2 use bogus items to detect malingering (exaggerating experiences such as symptoms of illness), social desirability (e.g., faking good), and careless responding (Baer et al., 1997; Baer, Kroll, Rinaldo, & Ballenger, 2000; Berry et al., 1992; Nichols et

al., 1986; Rogers et al., 2003). Although these items may be effective in high-stakes situations (e.g., court cases, job recruitment), they could be harmful in low-stakes situations (e.g., personality research involving college students). Specifically, bogus items do not have a clear option that participants know is the “right” response. When the purpose of a bogus item is unclear, participants might perceive the item differently than what the researcher intended (Meade & Craig, 2012; Schwarz, 1999). Participants might find bogus items to be funny (e.g., “I am paid biweekly by leprechauns;” Meade & Craig, 2012, p. 441), and not treat the survey seriously.

Instructed response items are as effective as bogus items, but do not have the ambiguity problem. They have a clear “correct” answer. Instructed response items ask participants to click a specific option (e.g., “Please select agree to this item;” McKay, Mussel, & Kaufman, 2014; Meade & Craig, 2012). McKay et al. (2014) examined similarities and differences in detecting careless responders with bogus items and instructed response items and found that the two types of items had a strong positive correlation with one another ($r > .50$). The two types of items are likely to be detecting the same type of response pattern and therefore it would be ideal to use the ones that do not have ambiguity in interpretation. These types of items, however, have not been used as often as bogus items have been for detecting careless responders. The effectiveness of instructed response items has not been investigated as thoroughly as have bogus items.

A final method is self-reported careless responding measures. In a number of studies using the MMPI-A and MMPI-2, Baer and colleagues (1997, 2000; Berry et al., 1992) found that more than 50 percent of participants in college samples and job applicants self-reported responding carelessly to at least one or more items, primarily in the middle and near the end of the survey. Meade and Craig (2012) tested measures in which participants self-reported the amount of attention and effort using single items. They found these items to be ineffective on their own and to be unrelated to other careless responding detection methods. They also included full scales to measure diligence and interest in survey participants. Although no cut point exists for what participants to exclude based on responses to these scales, they are easy to include and may provide information as to how much attentiveness and interest people used when completing a survey.

Post Hoc Data Screening Methods

There are a number of other indices to detect careless responders that are not as explicit in form. These post hoc methods involve screening the data for multivariate outliers (Ehlers, Greene-Shortridge, Weekley, & Zajack, 2009), comparing response patterns to semantic or psychometric antonyms and synonyms (Goldberg, 2000, cited in Johnson, 2005; Goldberg & Kilkowski, 1985; Maniaci & Rogge, 2014; Schinka, Kinder, & Kremer, 1997), or calculating response consistency by computing within-person split-half reliability correlations (i.e., even-odd consistency; Huang et al., 2012; Jackson 1976, as cited in

Johnson, 2005; Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012).

It is worth noting that the use of semantic antonym and synonym methods (comparing response patterns to semantically similar or dissimilar items) of screening for careless responding is likely to cause the exclusion of normal responders (Kurtz & Parrish, 2001) as the “failed” response patterns may just reflect personality differences (Johnson, 2005). Response patterns to semantic antonym and synonyms might also be influenced by vocabulary skills.

Participants who have limited vocabulary skills (e.g., English as a Second Language) might not understand the pair of antonyms or synonyms. Additionally, using Mahalanobis distance values (a statistical method to detect response patterns that deviate from “typical” responses) to screen for careless responders may be difficult when there are several items or may miss careless responders who respond in a “normal” manner. Moreover, this method is strongly positively correlated with bogus items ($r = .39$; Meade & Craig, 2012). In other words, using bogus items or instructed response items would be just as effective and easier in long surveys than would be the use of conducting multivariate outlier analysis.

There are two other post hoc methods that are useful and practical to detect careless responders. First, the researcher times participants while they complete a survey. Huang et al. (2012) found that this method was effective for detecting careless responders; people who spent less time completing a survey than would be reasonable would be considered careless responding. The researchers classified participants spending less than two seconds per item as

careless responders (see also Chiaburu, Huang, Hutchins, & Gardner, 2013). Curran et al. (2010) used a more liberal cut point of 5.5 seconds per item, but found similar rates of careless responding, which are discussed below. This method appears to produce results similar to other post hoc screening methods. Huang et al. (2012) found the results of this method to be moderately positively correlated with the psychometric antonym index and the even-odd consistency method. A second useful method is the LongString index (Behrend et al., 2011; Huang et al., 2012; Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012). The LongString index is calculated by determining the number of times a participant consecutively selects the same option. For the 300-item Revised Neuroticism-Extroversion-Openness Inventory (NEO-PI-R), Costa and McCrae (2008) developed certain cut points for each of the five response options (*strongly disagree-strongly agree*) that are used for the NEO-PI-R. A more common approach is by calculating the LongString index on a single webpage (Behrend et al., 2011; Maniaci & Rogge, 2014; Meade & Craig, 2012). Meade and Craig (2012) found the LongString index to detect a unique type of careless responder that the other methods did not detect well. Huang et al. (2012) also found this measure to have a weak positive correlation with the other indices that they used. They also found it to be effective for detecting careless responders that the other methods did not detect.

Summary of Screening Methods

The methods available to detect careless responders are diverse. In

experimental research, Oppenheimer et al.'s (2009) instructional manipulation checks are effective in making sure people read subtle changes in instructions. For survey research, Meade and Craig (2012) propose that researchers would benefit by using multiple methods. There are at least four methods that are effective and easy to use. First, because instructed response items are moderately to strongly correlated with multivariate outlier values, these items may be easier to use because Mahalanobis distance values may be difficult to calculate when there are a large number of items in a survey. Second, survey completion time, because of its reported relationship with other indices, is also an easy method to use that is undetectable by participants. Third, Meade and Craig (2012) found the LongString index detected a unique class of careless responders that most other methods did not catch. This method would be important to include because it is conceptually and practically different from the other methods. Last, self-reported measures at the end of the survey are simple and easy to use. These methods might also be useful for determining participants' effort when they complete a long survey.

Base Rates of Careless Responding

The methods described above to detect careless responders have detected varying frequencies of occurrence, depending on the method used. In lab settings, Oppenheimer et al.'s (2009) instructional manipulation checks flagged as low as 14 percent in a supervised setting and as high as 46 percent in

another study of unmotivated participants (nonstudents visiting the campus and receiving no compensation). Despite the loss of participants in these studies, the statistical power drastically increased based on the increase in strength of the effect sizes (see also Maniaci & Rogge, 2014). Kurtz and Parrish (2001) found 10.6 percent of their sample responded carelessly, but noted the method was most likely ineffective in detecting actual careless responders and likely flagged honest responders. Curran et al. (2010) found that rates of careless responding depended on the method used to detect careless responders. Response time flagged 6.5 percent of cases as careless responders, psychometric antonyms flagged 21.1 percent, the even-odd consistency flagged 50.5 percent, and the LongString index flagged 5.1 percent. Osborne and Blanchard (2010) used bogus items to identify careless responders in a sample of middle school students who completed a pre-test and post-test for an educational instructional intervention. They identified 40 percent of students to be careless responding on the pre-test and 29.5 percent to be careless responding on the post-test. Beach (1989) using bogus items found 4 percent of participants to be responding carelessly on a paper-and-pencil survey and 10 percent of participants to be responding carelessly when they completed the survey on a computer. Johnson (2005) found a base rate of 3.5 percent and noted it was comparable to rates in paper-and-pencil settings. It is important to note, however, that this lower percentage might be due to participants voluntarily completing the study for enjoyment and without any compensation. Recall, when compensation is

involved, participants can be more concerned about payment or credit rather than the meaning or implication of their participation (Meade & Craig, 2012). Ehlers et al. (2009) found roughly 5 percent of their sample of job applicants to be responding carelessly. Huang et al. (2012, Study 2) identified 5 percent of the sample of undergraduate students to be responding carelessly. Maniaci and Rogge (2014) identified 3-9 percent in three studies. Cho and Allen (2012) found 26 percent of parents recruited through Amazon MTurk had responded carelessly. McKay et al. (2014), using instructed response items, found 49 percent of undergraduate students and 20 percent from an MTurk sample who completed a survey online responded carelessly.

The various methods for detecting careless responders have been included in latent profile analyses to determine the different classes and rates of careless responders. Meade and Craig (2012) identified two classes of careless responders in their sample, which made up 11 percent of their entire sample. The first class, which was 9 percent of their sample, consisted of responders who spent less time on the survey, failed a large number of bogus items, did not answer items consistently based on the psychometric antonym and psychometric synonym methods, and had a higher Mahalanobis distance score. The second class consisted of 2 percent of the sample and was highlighted by high LongString index values, indicating these respondents clicked the same response option repeatedly. Maniaci and Rogge (2014) replicated these findings and found the same two classes of careless responders in their sample.

The percentages noted above are alarming given the evidence that even small percentages of careless responders in a dataset can decrease statistical power and increase the likelihood of a Type II error. Although there are methods available to identify and remove careless cases after data collection, a researcher's resources have already been spent. In online research, each participant receives payment, regardless of the quality of his or her data. Further, it costs time to screen the data to detect and remove careless responders from the dataset. Given the costs associated with detecting and removing careless responders' data, it is surprising that although methods to detect careless responders are becoming common in research, there are few methods currently available to deter people from careless responding in survey research.

Previous Attempts to Reduce Careless Responding

Previous research has examined the use of warnings or asking people to type their name at the end of a survey webpage in order to reduce careless responding. Huang et al. (2012) gave a stern warning that course credit would not be given to participants who were identified as responding carelessly. They found this to be an effective means to reduce careless responding. This warning, however, might be ineffective long-term if participants realize they will get credit regardless of how they completed the survey. It could also result in further participant misconduct towards a researcher. Meade and Craig (2012) did not include a statement worded as harshly. Rather, in one condition, they asked

people to type in their names on every webpage next to a statement about merging their data across web pages (identified condition). In another condition, they asked people to type their name next to a statement advising participants that responding honestly and carefully is in accordance with the university's honor policy (stern warning condition). These conditions were compared to a typical anonymous condition. The researchers found that when participants were anonymous in a typical online survey situation, they were significantly more likely to respond to bogus items incorrectly compared to the two identified conditions. People in the stern warning condition also failed significantly fewer bogus items than did those in the identified condition. There was a marginally significant difference on an interest scale that assessed how much interest was used when completing the scale with lower scores indicating higher interest in Meade and Craig's study. There was a lower mean score in the warning condition than there was in the anonymous and identified conditions, although it was possible that participants who identified themselves might have been more motivated to "fake good." The researchers did not find a difference in social desirability scores in their two identified conditions compared to an anonymous condition. They concluded that having participants identify themselves could be effective in reducing careless responding.

When participants complete sensitive surveys, differences between anonymous and nonanonymous situations have been observed. Specifically, when participants identify themselves completing surveys asking about sensitive

issues (e.g., sexual behaviors, cheating), they might engage in socially desirable responding. Ong and Weiss (2000) found people were more likely to report engaging in risky behaviors and more likely to report having cheated on a test at some point when the survey was anonymous compared to when participants provided their name and were told the survey would remain confidential. Booth-Kewley, Larson, and Miyoshi (2007) found that people reported consuming more alcohol and engaging in risky sexual behaviors when the survey was online compared to people completing a paper-and-pencil survey because of the greater anonymity associated with online surveys. Joinson (1999) found that when participants were completing a survey anonymously online, participants were least likely to respond in a socially desirable manner. The highest socially desirable scores were from participants completing a paper-and-pencil survey and who were not anonymous. Richman, Kiesler, Weisband, and Drasgow (1999), in a large meta-analysis, compared socially desirable responding on paper-and-pencil tests and in face-to-face interview. They found that people were less likely to respond with socially desirable answers/responses on paper-and-pencil tests. They found, however, that participants responded more socially desirable on computers, but this was reduced when the participant was alone and responding anonymously.

Not all studies have found differences in behaviors reported in anonymous versus nonanonymous situations; nonetheless, it may be ideal to retain an anonymous survey environment for participants, especially if it is unnecessary to

have participants identify themselves. Furthermore, because anonymity is preferred over confidential surveys in terms of ethics, it is better to focus on ways to reduce careless responding that also retain anonymity. This approach might be most important for doing research with sensitive populations (convicts, psychiatric patients, children, etc.) that do not allow any identifying information to be collected. It also would be beneficial to develop a method that would meet the standards of even the most stringent IRBs. This method, which overlaps with Meade and Craig's (2012) stern warning condition, involves the use of an honor code presented at the beginning of a study. Additionally, pictures of watchful eyes might have similar effects to honor codes.

Honor Codes and Watchful Eyes: Their Effects on Behavior

There are two methods that may be effective in reducing careless responding that could also retain anonymity. These methods involve increasing self-awareness and moral behavior through the use of honor codes and watchful eyes. Moral identity is an integral component to one's identity (Aquino & Reed, 2002). People, however, may engage in dishonest behaviors in contexts such as dim lighting (Zhong, Bohns, & Gino, 2010), when cognitive resources are depleted (Gino, Schweitzer, Mead, & Ariely, 2011), or in the presence of abundant amounts of money (Gino & Pierce, 2009). Theories on self-awareness—the act of shifting focus to internal psychological processes—have been identified as a method to reduce dishonest behavior (Mažar & Ariely, 2006).

Self-awareness can be induced through the presence of a mirror (Beaman, Klentz, Diener, & Svanum, 1979; Dijksterhuis & Van Knippenberg, 2000; Vincent, Emich, & Goncalo, 2013), religious priming (Gervais & Norenzayan, 2012; Mažar, Amir, & Ariely, 2008), honor codes (Mažar et al., 2008; Shu, Gino, & Bazerman, 2011; Shu, Mažar, Gino, Ariely, & Bazerman, 2012), or watchful eyes (Bateson, Nettle, & Roberts, 2006; Ernest-Jones, Nettle, & Bateson, 2011). The current study focused the use of two of these methods – honor codes and watchful eyes.

Honor Codes

Honor codes are typically used in academic environments to reduce cheating and plagiarism. McCabe, Trevino, and Butterfield (2002; McCabe & Trevino, 1993, 1997) reported that one method typically utilized by universities to reduce cheating is by having students sign an honor statement or a pledge. Universities that typically use these approaches have lower rates of cheating than do universities that do not use them. In their analyses of dishonest behaviors, Mažar et al. (2008) observed that it was not the honor code itself per se that reduced dishonest reporting, but it was the reminder to behave morally cued by the honor code. In another study, peoples' self-awareness was increased after being instructed to write as many of the 10 commandments as they could remember in two minutes. This method also led to a reduction in dishonesty. Shu et al. (2011) found that an honor code significantly reduced dishonest behavior. In fact, dishonesty almost completely disappeared. Similarly,

Shu et al. (2012) utilized a method to increase honest reporting. They asked participants to sign a document before reporting travel mileage for reimbursement at the beginning, rather than the end, of a tax reimbursement document. This simple act subsequently decreased the amount of mileage reported, perhaps because participants were more self-aware of their morality if they signed the document before completing it.

In the cases above, honor codes demonstrated strong effects on moral behavior. This effect is likely to generalize to students completing surveys. Having a person read and acknowledge an honor code before completing a survey is likely to reduce careless responding.

Watchful Eyes

Another useful method to increase self-awareness and reduce dishonest behavior is with watchful eyes. Bateson et al. (2006) found that hanging a picture of a pair of eyes above a donation box led people to donate more money than when there was a picture of flowers above a donation box. This change led to an increase in self-awareness of people's behavior and resulted in more money donated. Ernest-Jones et al. (2011) replicated this same effect in a cafeteria setting. They found that when a picture of a pair of eyes was hung on the wall, people were more likely to throw their trash away after eating than were people when there was a picture of flowers on the wall.

Honor Codes and Watchful Eyes

In order to examine the effects of these two methods, an honor code can

be presented alone or in combination with a picture of a pair of eyes. These methods used in combination could likely have a stronger effect on participant behavior and increased self-awareness. One problem, however, with increased self-awareness is people might distort responses in a socially desirable manner. Gervais and Norenzayan (2012) found increased self-awareness led to higher scores on a measure of social desirability. This problem, however, can be addressed. Dwight and Donovan (2003) found that asking or warning participants to respond truthfully to a survey led to lower scores on a social desirability measure compared to when people were not asked or warned to respond truthfully. Therefore, including a sentence about responding truthfully to items may neutralize possible negative effects to responses for sensitive survey items. The present study sought to address the effectiveness of an honor code with or without the presentation of a picture of watchful eyes at the start of a survey. Additionally, highly sensitive survey measures were used to determine if socially desirable responding might occur.

The Present Study

The purpose of the present study was to examine the effectiveness of increasing people's self-awareness with the use of an honor code and the presentation of a picture of eyes before completing a survey in order to reduce careless responding. Participants provided consent and then were randomly assigned to one of five conditions in which the presentation of an honor code

and/or a pair of eyes differed. The five conditions consisted of a read-only condition, type-only condition, read-and-eyes condition, or type-and-eyes condition, or a control condition (no honor code). These conditions are more thoroughly explained in Chapter 2. Participants then completed a number of self-report personality and social perception measures split across three webpages, which had four instructed response items strategically placed on each of the three pages.

The honor codes were displayed on their own webpage after the informed consent to increase the likelihood participants would read it. Cassileth, Zupkis, Sutton-Smith, and March (1980) found that 40 percent of people report that they do not always read the informed consent “carefully.” People view the informed consent as a method to protect the physician or the researcher rather than as a method to protect the patient or research participant even for high stakes medical procedures. This tendency means that placing an honor code within the informed consent may result in a reader skipping the honor code altogether, so it will not have an effect on response quality. It could also indicate that people would skip over the honor code or not read it carefully in the read-only condition.

Three indices for detecting careless responding were used in the present study. These indices were instructed response items, the average LongString index (Meade & Craig, 2012), and time to complete the survey (Huang et al., 2012). These methods allowed me to test rates of careless responding among the five conditions. Participants also completed Meade and Craig’s (2012)

participant engagement scale at the end of the survey, which measures diligence and interest when completing the surveys.

Based on these conditions and the information presented, I tested two hypotheses and six research questions:

Hypothesis 1

I hypothesized that careless responding would be reduced in the honor code conditions that involved typing an honor code and/or in conjunction with viewing a picture of a pair of eyes (type-only, read-and-eyes, and type-and-eyes) compared to the control condition or the read-only condition.

Hypothesis 1a. I predicted that there would be fewer instructed response items failed in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition.

Hypothesis 1b. I predicted that there would be a smaller average LongString index in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition.

Hypothesis 1c. I predicted that there would be more time spent completing the survey in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition.

Hypothesis 2

I hypothesized that there would be higher scores on the diligence scale and the interest scale in the honor code conditions that involved typing an honor code and/or in conjunction with viewing a picture of a pair of eyes (type-only,

read-and-eyes, and type-and-eyes) compared to the control condition or the read-only condition.

Hypothesis 2a. I predicted that there would be significantly higher scores on the diligence scale in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition.

Hypothesis 2b. I predicted that there would be significantly higher scores on the interest scale in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition.

Research Question #1

Among the three conditions that I predicted to be the most effective at reducing careless responding (type-only, read-and-eyes, type-and-eyes), would one of the three be more effective than the others at reducing careless responding?

If there were significant group differences among these three conditions, I conducted 2 (Honor Code: read, type) × 2 (Eyes: no eyes, eyes) between-person ANOVAs. These post hoc analyses allowed me to detect whether the presence of eyes with the different honor code conditions or the honor code alone had a greater influence reducing careless responding.

Research Question #2

Would careless responding be significantly reduced in the read-only

condition compared to the control condition?

Research Question #3

Previous research has also found that people self-report greater careless responding towards the middle and end of the survey (Berry et al., 1992).

Because four instructed responding items were placed on each of the three webpages, I was able to test whether differences in careless responding exist across the survey. Using a within-person design, I examined rates of careless responding over the span of the survey.

Does careless responding increase in the middle and end of the survey or does it remain consistent throughout the survey?

Research Question #4

In this study, I have chosen to use personality measures containing items that people might be inclined to “fake good” when self-awareness is heightened. People may rate themselves higher on positive attributes (e.g., the honesty/humility factor from the HEXACO structure of personality) or rate themselves lower on negative attributes (e.g., Machiavellianism, risky behaviors). Because ratings to these personality scales may be susceptible to honor codes or watchful eyes, I examined if there were differences among the five conditions to ensure these methods do not lead to socially desirable responding. I also included a sentence asking participants to answer truthfully. Dwight and Donovan (2003) found that either asking or warning people not to fake their responses was successful in reducing faking.

Would the conditions predicted to reduce careless responding (type-only condition, read-and-eyes condition, type-and-eyes condition) lead people to “inflate” their scores to certain personality measures that may be susceptible to heightened self-awareness? For example, would people report higher scores on honesty/humility and higher scores on a measure of social desirability in the type-only, read-and-eyes, or type-and-eyes conditions compared to people in the control condition?

Research Question #5

Because the personality measures used in this study reflected traits that could relate to careless responding, it would be practical to calculate correlations among the careless responding indices and the personality measures. Calculating these correlations, however, should be contingent on the results of Research Question #4. If scores on the personality measures are not influenced by the conditions predicted to increase self-awareness, their relationships might shed light on dispositional traits that relate to careless responding. It would be likely that “positive” personality traits (e.g., honesty/humility, conscientiousness) would negatively relate to careless responding (i.e., higher scores on these personality measures would predict less careless responding). “Negative” personality traits (e.g., Machiavellianism, risky behaviors) might also positively relate to careless responding. These negative traits would likely present evidence of higher scores on these personality measures predicting greater careless responding.

If the conditions predicted to increase self-awareness do not influence/change scores on the personality measures, what personality traits would be related to the careless responding indices?

Research Question #6

Last, it would be important to investigate the impact of the platform (i.e., tablet, computer, cellular phone) that people use on careless responding. That is, if a participant uses a tablet or a cellular phone to complete the survey and the website is not compatible with those devices, participants might be inclined to respond carelessly. To determine this, I asked participants what device they used to complete the survey and if they experienced technological issues while completing the study. I used this information to determine how it relates to careless responding.

How does the platform (i.e., computer, tablet, cellular phone) of completing a study relate to careless responding?

CHAPTER TWO

METHOD

Participants and Recruitment

I conducted a power analysis using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) for conducting a one-way ANOVA with five conditions. I set the effect size at .25, error rate at .05, and power at .80. I needed to collect 200 participants to detect differences among conditions. Participants were required to be over the age of 18.

Undergraduate college students ($N = 349$) from a public university in Southern California signed up for the study through the psychology department's SONA system research participant management online software. Participants were compensated with three extra credit points. Data was collected during the last two weeks of a 10-week quarter. This period of time is thought to have higher rates of careless responding because respondents are hurriedly trying to complete studies for extra credit before the end of the quarter/semester (Huang et al., 2012). After screening for and deleting duplicate cases and cases with extensive missing data, I had 305 (87.39%) useable cases. The criterion for deleting cases is discussed in Chapter 3. Participants were between the ages of 18-70 ($M_{age} = 22.81$, $SD_{age} = 5.41$, $Mdn_{age} = 22$). There were 266 women (87.21%) and 39 men (12.62%). Participant's self-reported ethnicity was: 181 (59.3%) reported being Hispanic American/Latino(a), 53 (17.4%) reported being

White/Caucasian/European American, 30 (9.8%) reported being Asian/Asian American/Pacific Islander, 20 (6.6%) reported being African American/Black, 13 (4.3%) reported being Biracial, 6 (2.0%) reported “other,” and 2 (0.7%) reported being Middle Eastern/Arab.

Materials

For the honor code conditions, I used an honor code created for this study and a picture of a pair of eyes from Bateson et al. (2006). There were three methods to detect careless responding, including: (a) instructed response items, (b) time spent completing the survey, and (c) average LongString index (Huang et al., 2012; Meade & Craig, 2012). Participants also completed a participant engagement measure at the end of the survey (Meade & Craig, 2012).

Participants completed a number of personality measures, including: (a) the 100-item HEXACO (Lee & Ashton, 2004), (b) the 20-item Machiavellianism Scale (MACH-IV; Christie & Geis, 1970), (c) the 40-item Domain-Specific Risk-Taking Scale (DOSPERT; Weber, Blais, & Betz, 2002), (d) the 15-item Right Wing Authoritarianism scale (RWA; Zakrisson, 2005), (e) the 16-item Social Dominance Orientation scale (SDO; Pratto, Sidanius, Stallworth, & Malle, 1994), (f) a 10-item Self-Esteem scale (Rosenberg, 1965), (g) the 40-item Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991), and (h) the 64-item Self-Reported Psychopathy-III scale (SRP-III; Williams, Nathanson, & Paulhus, 2003).

Measures to Reduce Careless Responding

Participants in all of the conditions, except for the control condition, read or typed a four-sentence honor code before completing the survey. In the eyes conditions, participants were also exposed to a picture of a pair of eyes from Bateson et al. (2006) with the honor code. The honor code and picture of eyes are in Appendix B.

Measures to Detect Careless Responding

There were three methods to detect careless responding and some secondary measures to compare participant engagement among conditions.

The first method to detect careless responding was the inclusion of instructed response items mixed throughout the survey (Meade & Craig, 2012). They were in the same position in the survey for all participants. See Appendix C for samples of the instructed response items. The second method was the use of the average LongString index (Meade & Craig, 2012). This index was calculated by determining the longest consecutive string of the same number selected on a given webpage. For example, a participant clicking “3” 10 times in a row on a webpage—and never another instance with 10 or more of the same option selected consecutively on the same webpage—would receive a score of 10 on this index. This score was calculated by averaging the three webpages’ LongString index scores. This index was calculated using a Visual Basic Application for Microsoft Excel. Last, participants were timed while completing the survey questions. Typically, participants are timed throughout the entire study

(Huang et al., 2012). In the case of this study, calculating the time spent on the entire survey would introduce a confound in the two conditions that involved typing the honor code. It is likely that participants in these two conditions spent more time on the honor code page because they had to type the honor code, whereas participants in the other three conditions did not have to type the honor code. Additionally, if people reported experience technological issues, which this question was on the demographics page, they were also asked to type out what issue they experienced. Therefore, I started timing participants directly after the honor code page and stopped before the demographic page across conditions. The survey time variable reflected time spent completing the three primary webpages.

The secondary measures I used are a set of participant engagement measures developed by Meade and Craig (2012). These items followed the primary measures and demographics page, but were completed before the debriefing statement. There was also a single yes/no item (UseMe) asking if participants believe their data should be used by the researchers in the analyses (Meade & Craig, 2012). See Appendix D for the participant engagement measures.

Personality Measures

Participants completed a number of self-reported personality measures. All of the measures were completed using a 7-point Likert scale with endpoints varying between the measures; end points are provided with their respective

measures. The personality measures are explained in the order in which participants completed them in the study. See Table 1 for Cronbach's α and correlations among the personality measures for the full sample ($N = 305$).

HEXACO-100. Lee and Ashton's (2004) 100-item scale was used to measure the six-factor model of personality. The six personality factors in the HEXACO model are: (a) Honesty/Humility, (b) Emotionality, (c) Extraversion, (d) Agreeableness, (e) Conscientiousness, and (f) Openness to Experience. There are 16 items for each of the six factors with an additional four items making up an Altruism scale. The 16 items for each factor can also make up four subscales. I did not use the four subscales within each factor or the 4-item altruism scale because they were beyond the scope of the present study. Therefore, no analyses were reported based on these subscales. Participants completed the HEXACO items using the end points *strongly disagree* and *strongly agree*. Lee and Ashton (2004) provided evidence of internal consistency reliability for the factor scores (Cronbach's $\alpha = .89-.92$) and convergent validity with three Big Five personality scales and a measure of Psychopathy (with the Honesty/Humility factor). All correlations for convergent validity were greater than the absolute value of .68. The items were averaged to create a final personality score for each personality factor. See Appendix E for the HEXACO-100 items.

Machiavellianism (MACH-IV). Christie and Geis's (1970) 20-item MACH-IV scale was used to measure Machiavellianism type personality. There are three subscales in the MACH-IV (Views, Tactics, and Morality). I did not use the

subscales in the current study because they were beyond the scope of the present study. Items were rated using the endpoints *strongly disagree* and *strongly agree*. Zook (1985) stated that the MACH-IV has strong internal consistency scores (Cronbach's $\alpha = .70-.80$) and had stable test-retest scores over a six-week period (.76). The MACH-IV is typically tested with the personality traits Narcissism and Psychopathy, which the three traits make up the "dark triad" (Paulhus & Williams, 2002). The MACH-IV has been tested with these measures and has been tested for convergent validity with these other personality traits and the HEXACO's Honesty/Humility factor (Lee & Ashton, 2005; Paulhus & Williams, 2002). The items were averaged to create a final score. See Appendix F for the MACH-IV items.

Domain-Specific Risk-Taking (DOSPERT). Weber et al.'s (2002) 40-item domain specific risk taking scale was used. The scale can be used with one of two instructions, which allow participants to report how likely they are to engage in a number of risky behaviors or how much they perceive a situation to be risky. I assessed participant's likelihood of engaging in risky behaviors. The scale assesses five domains of risk taking, including: (a) Finance, (b) Health/Safety, (c) Recreation, (d) Ethics, and (e) Social. In the current study, Items were rated using the endpoints *very unlikely* and *very likely*. Weber et al. (2002) provided evidence of internal consistency reliability for the factor scores (Cronbach's $\alpha = .69-.83$). They also provided evidence of adequate test-retest reliability over a one-month period (.42 for the financial subscale to .67 for the ethics subscale),

convergent validity (with a risk-attitudes scale and sensation-seeking scale), and discriminant validity (with an intolerance for ambiguity scale). In a third study, they reduced the number of items in the scale to 40 (eight per subscale), which led to higher internal consistency (Cronbach's $\alpha = .74-.84$). They also found that men reported significantly higher scores on the risky behavior version than did women. The items were averaged to create a final score. See Appendix G for the DOSPERT items.

Right Wing Authoritarianism (RWA). Zakrisson's (2005) 15-item measure was used to assess right-wing authoritarianism. Participants rated the items using the endpoints *strongly disagree* and *strongly agree*. The 15-item version is a short form version of the full 30-item scale developed by Altemeyer (1998). The short form was created by removing items with the lowest contribution to the overall scale reliability until 15 items remained. The final Cronbach's α reported by Zakrisson (2005) was .72 and the scale had discriminant validity with the SDO scale. See Appendix H for the RWA items.

Social Dominance Orientation (SDO). Pratto et al.'s (1994) 16-item measure was used to assess social dominance orientation. Items were rated using the endpoints *strongly disagree* and *strongly agree*. Pratto et al. (1994) examined the reliability of the scale and tested for predictive, discriminant, and convergent validity with 13 different samples of college students from three different universities. The SDO had an average internal consistency of .83 for the 13 samples. For predictive validity, Pratto and colleagues found that SDO

predicted: (a) differences between men and women, (b) seeking certain hierarchical roles, (c) agreement with ideologies of hierarchical-legitimizing myths, and (d) agreement with policy attitudes. For discriminant validity, they found that the SDO scale was correlated with political conservatism and an RWA scale, but the SDO predicted policy attitudes beyond political conservatism and the RWA scale. They tested for convergent validity with an empathy, altruism, and communality scale and a personality scale assessing tolerance (i.e., accepting other people's beliefs and values). The items were averaged to create a final score. See Appendix I for the SDO items.

Self-Esteem. Rosenberg's (1965) 10-item self-esteem scale was used to measure participants' self-esteem. Items were rated using the endpoints *strongly disagree* and *strongly agree*. Heatherton and Wyland (2003) reported that the Rosenberg self-esteem scale is the most commonly used scale for measuring self-esteem. They reported it has an internal consistency of .92. Schmitt and Allik (2005) across 53 cultures and in 28 languages found a reliability coefficient of .88 in the U.S. sample (range was from .61 in Tanzania to .90 in the U.K. and Israel). Schmitt and Allik also demonstrated convergent validity (extraversion, neuroticism, and model of self scale) and discriminant validity (openness and model of other scale). The items were averaged to create a final score. See Appendix J for the self-esteem items.

Balanced Inventory of Desirable Responding (BIDR). Paulhus's (1991) 40-item social desirability measure was used. There are two subscales within this

measure, which are Self-Deceptive Enhancement (SDE) and Impression Management (IM). Participants rated the items using the end points *not true* and *very true*. Paulhus reported internal consistencies of .68 to .80 for the SDE subscale and .75 to .86 for the IM subscale. Test-retest reliability over a five-week period was .69 for the SDE and .65 for the IM. The measure has present concurrent validity with the commonly used Marlowe-Crowne social desirability scale. Paulhus has also demonstrated convergent validity with the SDE scale and other measures such as repressive styles, defense mechanisms, and coping styles. Convergent validity for the IM scale has been demonstrated with measures such as lie scales, agreeableness, and conscientiousness. Discriminant validity was demonstrated with the formation of two different factors through factor analysis. The items for each of the subscales were averaged to create two final scores. See Appendix K for the BIDR items.

Self-Reported Psychopathy-III (SRP-III). Williams et al.'s (2003) 64-item scale was used to measure self-reported sub-clinical psychopathy. There are four subscales to the SRP-III, including: (a) Interpersonal Manipulation, (b) Callous Affect, (c) Erratic Life Style, (d) Anti-Social Lifestyle. All of the items can also be treated as one scale. I used one scale because the subscales were beyond the scope of the present study. Items were rated using the end points *strongly disagree* and *strongly agree*. Williams et al. found an alpha reliability of .88 for the full scale. The SRP-III is typically included with the personality traits Narcissism and Machiavellianism, which together the three traits make up the

“dark triad” (Paulhus & Williams, 2002). Williams et al. (2003) found that the SRP-III was positively correlated with the other two dark triad traits and negatively correlated with agreeableness and conscientiousness. They also demonstrated that the SRP-III was strongly positively correlated with other measures of psychopathy, but the relationships were not strong enough to indicate they were measuring the same traits exactly. The items were averaged to create a final score. See Appendix L for the items in the SRP-III.

Procedure

Participants signed up for the study using the CSUSB Psychology Department’s SONA system. They were then redirected to the *Qualtrics* online survey system to complete the survey. The survey measures did not utilize a response web design (i.e., a website that adjusts to the screen size of different devices such as a tablet or cellular phone). Upon accessing the survey, participants were presented with an informed consent statement stating that they would complete a number of personality measures assessing different personality characteristics. See Appendix A for the informed consent. Following the informed consent webpage, people were randomly assigned to one of five honor code conditions. These conditions were: (a) read-only, (b) type-only, (c) read-and-eyes, (d) type-and-eyes, and (e) control condition (no honor code). In all of the conditions, participants read one sentence with general instructions for the study. The information below that sentence varied depending on the

condition. In the read-only condition, participants were instructed to read an honor code on the screen. In the type-only condition, participants were instructed to type the honor code into a box provided on the webpage. In the read-and-eyes condition, participants were instructed to read the honor code and above the honor code was a picture of a pair of watchful eyes. A picture of a pair of eyes has been found to increase self-awareness in a number of real-world settings (Bateson et al., 2006; Ernest-Jones et al., 2011). In the type-and-eyes condition, participants were required to type the honor code into a box provided on the webpage and above the honor code was a picture of a pair of watchful eyes. In the control condition (no honor code), there was no honor code or watchful eyes on the webpage. To restrict participants from simply copy/pasting the honor code in the two type conditions, the honor code was presented as an image rather than as text. See Figure 1 in Appendix P for a screenshot of the type-and-eyes honor code webpage. Following the honor code page, participants were then instructed to complete the personality measures.

The personality measures were divided up among three primary webpages with 100 survey items on the first page, 101 survey items on the second page, and 104 survey items on the third page. The survey items were randomized prior to data collection and were in the same order for all of the participants. On each of the webpages, instructed response items were placed in the same position for each participant. On the first webpage, participants completed the 100-item HEXACO scale (Lee & Ashton, 2004). There were four

instructed response items placed as the 20th, 40th, 60th, and 80th items on the first webpage. On the second webpage, participants completed the 20-item MACH-IV scale (Christie & Geis, 1970), the 40-item DOSPERT scale (Weber et al., 2002), the 15-item RWA scale (Zakrisson, 2005), the 16-item SDO scale (Pratto et al., 1994), and the 10-item self-esteem scale (Rosenberg, 1965). There were four instructed response items placed as the 30th, 50th, 70th, and 90th items on the second webpage. On the third webpage, participants completed the 40-item BIDR (Paulhus, 1991) and the 64-item SRP-III (Williams et al., 2003). There were four instructed response items placed as the 25th, 45th, 65th, and 85th items on the third webpage. The webpages had 104, 105, and 108 items in their respective presentation order to participants. On each of the personality survey pages, time spent was recorded by the *Qualtrics* system without participants' awareness.

On the fourth webpage, participants completed the demographic items (see Appendix M) and participant engagement measures (Meade & Craig, 2012). On the final page, participants viewed the debriefing statement, which explained the purpose of the experiment and thanked the participants for their time.

CHAPTER THREE

RESULTS

Sample

Handling of Duplicate Cases and Missing Data

Of the 349 participants who signed up to complete the study, three people declined to participate. These three people were not included in the analyses. There were 13 participants with duplicate cases (26 cases total). Duplicate cases were identified with their SONA ID number. In order to handle duplicate cases, I deleted the case with more missing data. In other words, cases were removed if one of the two duplicate cases did not complete at least one of the webpages. Each of the 13 participants with duplicate cases did not complete at least one of the webpages on one of their two cases. Therefore, the case with more missing data was removed. One of the participants did not complete any of the webpages for either of their two cases; both cases were removed. In total, 14 of the 26 duplicate cases were removed and not included in the analyses.

Next, cases with extensive missing data were removed. Because careless responding was determined based on the instructed response items (four per webpage), the average LongString index, and survey time across the three webpages, it was important for participants to have at least partial data on each of the three webpages. If a participant did not complete a webpage, the number of instructed response items would be out of 4 or 8 rather than out of 12;

participants' time spent completing the webpage would be shorter or not provided by the website, and their LongString index would be averaged across two webpages rather than three webpages. Therefore, extensive missing data was defined as cases not completing all of the survey items on at least one of the webpages. There were five people who completed the first and second webpages, but not the third webpage. There were 11 cases that did not complete any of the survey measures on any of the three webpages. There were 11 people who completed the first webpage, but not the second and third webpages. As a result, there were 27 cases that were removed because of extensive missing data and not included in the analyses.

Of the 349 people who signed up for the study, there were 44 (12.61%) people who were removed. It was important to determine if there was greater attrition in one of the experimental conditions compared to another condition. The number of people who dropped out in each condition was: (a) seven people in the control condition, (b) 10 people in the read-only condition, (c) six people in the type-only condition, (d) seven people in the read-and-eyes condition, and (e) nine people in the type-and-eyes condition. Dropout rates in each condition were too small to provide sufficient power for conducting a statistical analysis to compare differences among conditions. The rates, however, appeared similar across the five conditions. It is worth noting that the 15 people who were removed from the type-only and type-and-eyes conditions typed out the honor code in full before dropping out of the study. There is no explanation for why

people typed the honor code, but did not complete any of the measures ($n = 5$) or a small portion of the measures ($n = 10$). After screening for duplicate cases and cases with missing data, there were 305 useable cases for the analyses (87.39%).

Test of Hypotheses and Research Questions

Reduction in Careless Responding among Conditions

Hypotheses 1a, 1b, and 1c and Research Questions #1 and #2. Three one-way fixed effects ANOVAs were conducted to test Hypotheses 1a, 1b, and 1c and Research Questions #1 and #2. For Hypothesis 1, I predicted that careless responding would be reduced in the honor code conditions that involved typing an honor code and/or in conjunction with viewing a picture of a pair of eyes (type-only, read-and-eyes, and type-and-eyes) compared to the control condition or the read-only condition. The independent variable was the honor code condition, which had five levels (no honor code, read-only, type-only, read-and-eyes, and type-and-eyes). The one-way ANOVAs were conducted on the instructed response items (range 0-12; H1a), average LongString index (H1b), and survey time (H1c).

First, I examined the data for normality and homogeneity of variance. Normality was assessed for each variable by calculating a z score for skewness for each honor code condition. A significantly skewed variable was defined as having a z score for skewness ± 3.3 . In all of the five conditions, the instructed

response item variable ($z_s = 5.62-12.10$), average LongString index ($z_s = 3.81-22.86$), and survey time variable ($z_s = 17.41-24.97$) were significantly positively skewed. These positively skewed variables indicated that most people had lower scores on these variables and that there were few people who had high scores on these variables. The assumption for homogeneity of variance was violated for the instructed response items (Levene's test = 12.01, $p < .001$) and average LongString index (Levene's test = 7.22, $p < .001$) ANOVAs. The ratio of the condition with the largest variance to the condition with the smallest variance was also greater than 4:1. Because of the nonnormal distribution and violation of homogeneity of variance, I followed recommendations by Erceg-Hurn and Mirosevich (2008) and used bootstrapping techniques, which are robust for analyzing nonnormal data. Additionally, it is recommended (see Ratcliff, 1993; van Zandt, 2002; Yan & Tourangeau, 2008) that trimmed means, median scores, or harmonic means be used for response time or survey time in analyses compared to using an uncorrected mean score. Yan and Tourangeau (2008) found that these three procedures were highly correlated ($r > .90$). Because survey time was highly positively skewed, I trimmed 10 percent within each condition from only the right tail of the distribution based on Ratcliff's (1993) recommendations. There were six cases removed from the control condition, seven cases from the read-only condition, six cases from the type-only condition, six cases from the read-and-eyes condition, and five cases from the type-and-eyes condition. I chose not to transform the data so the results of the current

study could be compared to previous studies using similar measures for detecting careless responders (see Meade & Craig, 2012). Meade and Craig reported uncorrected mean scores for the instructed response items and average LongString index. For consistency across the three DVs in the current study, I conducted the one-way ANOVAs with bootstrapped 95 percent bias corrected confidence intervals (1,000 samples). For significant post hoc pairwise comparisons, Tukey's Honestly Significant Differences (HSD) tests were used. See Table 2 in Appendix O for F -values, partial η^2 s, means, standard deviations, and the 95 percent bias corrected confidence intervals for each condition for each ANOVA.

There was a significant mean difference on the instructed response items based on honor code condition (see Figure 2 in Appendix P). There were significantly more instructed response items failed in the read-only condition than there were in the control condition. There was a marginally significant difference between the read-only condition and the read-and-eyes condition ($p = .070$) and the read-only condition and the type-and-eyes condition ($p = .055$). There were more instructed response items failed in read-only condition compared to in the read-and-eyes condition and type-and-eyes condition. The remaining pairwise comparisons for the instructed response items were nonsignificant. There was a significant mean difference on the average LongString index based on honor code condition (see Figure 3 in Appendix P). There was a significantly greater average LongString index in the read-only condition than there was in the control

condition. The remaining pairwise comparisons for the average LongString index were nonsignificant. There was a nonsignificant mean difference on survey time based on honor code condition (see Figure 4 in Appendix P). As a result, hypotheses 1a, 1b, and 1c were not supported and the pattern of findings was inconsistent with the predictions.

For Research Question #1, I examined differences among the three honor code conditions that I predicted to be the most effective at reducing careless responding (type-only, read-and-eyes, and type-and-eyes). To answer this research question, I examined the pairwise comparisons in the three ANOVAs reported above among the type-only condition, the read-and-eyes condition, and the type-and-eyes condition. These three conditions did not significantly differ from one another. The scores in these conditions also did not significantly differ from the control condition. Therefore, I did not conduct the post hoc 2 (Honor Code: read, type) \times 2 (Eyes: no eyes, eyes) between-person ANOVAs.

For Research Question #2, I examined if careless responding would be reduced in the read-only condition compared to in the control condition. To address this research question, I determined if there were significant post hoc pairwise differences between the control condition and the read-only condition with the three ANOVAs reported above. There were significantly more instructed response items failed in the read-only condition compared to in the control condition. There was also a significantly greater average LongString index in the read-only condition compared to in the control condition. Based on the results,

there was an increased in careless responding in the read-only condition compared to the control condition rather than a decrease in careless responding in the read-only condition.

Hypothesis 2. Two one-way fixed effects ANOVAs were conducted to test Hypothesis 2. The independent variable was the honor code condition, which had five levels (control condition, read-only, type-only, read-and-eyes, and type-and-eyes). The two one-way ANOVAs were conducted on the diligence scale and the interest scale. For Hypothesis 2, I predicted that there would be significantly higher scores on the diligence scale (H2a) and interest scale (H2b) in the honor code conditions that involved typing an honor code and/or in conjunction with viewing a picture of a pair of eyes (type-only, read-and-eyes, and type-and-eyes) compared to the control condition or the read-only condition.

I examined the participant engagement subscales for normality by calculating a z score for skewness for each honor code condition. A significantly skewed variable was defined as having a z score for skewness ± 3.3 . Scores on the diligence subscale were significantly negatively skewed in four of the five conditions with scores not being skewed in the type-only condition. These negatively skewed z scores indicated that most people had higher scores on the diligence subscale variables and that there were few people with lower scores on diligence subscale. The assumption for homogeneity of variance was not violated for either variable. I did not transform the scores on the diligence scale, but conducted two one-way ANOVAs with bootstrapped 95 percent bias corrected

confidence intervals (1,000 samples) for consistency with the careless responding measures. For significant post hoc pairwise comparisons, Tukey's HSD tests were used. See Table 2 in Appendix O for F -values, partial η^2 s, means, standard deviations, and the 95 percent bias corrected confidence intervals for each condition for each ANOVA. There was a nonsignificant effect on the diligence scale (see Figure 5 in Appendix P) and interest scale (see Figure 6 in Appendix P) based on honor code condition. Hypotheses 2a and 2b were not supported.

Research Question #3. Three repeated-measures ANOVAs were conducted to test Research Question #3. The independent variable was the webpages, which had three levels (page 1, 2, and 3). The repeated-measures ANOVAs were conducted on the instructed response items on each page (0-4), the maximum LongString index (i.e., the longest consecutive string of values on each of the three webpages), and page time (time spent on each webpage). For the time spent on webpage analysis, all participants were included and means were not trimmed. See Table 3 in Appendix O for the F -values, partial η^2 s, means, and standard deviations for each webpage. For the instructed response item ANOVA, sphericity was not violated. For the maximum LongString and page time, sphericity was violated based on Mauchly's test of sphericity. Therefore, the Greenhouse-Geisser corrected F -values are reported. There was a nonsignificant effect for webpages on the instructed response items and page time. There was a significant effect for webpages on the maximum LongString

index. Specifically, there was a significantly greater maximum LongString index on page 2 compared to on page 1 and there was a significantly greater maximum LongString index on page 3 compared to on page 1. There was a nonsignificant difference on the maximum LongString index between page 2 and page 3.

Because the standard deviations were large for the maximum LongString index and page time variable, I excluded outliers on each of the within-person variables and conducted the same three repeated-measures ANOVAs. Outlier cases were defined as having *z* scores on a webpage on each dependent variable that was ± 3.3 . For the instructed response items, there were seven outliers on page 1 (all *z*s = 3.97), 11 outliers on page 2 (all *z*s = 3.63), and 13 outliers on page 3 (all *z*s = 3.62). There were 19 unique outliers—some of the cases were outliers on more than one webpage—excluded for the instructed response item analysis. For the maximum LongString index, there were three outliers on page 1 (*z*s = 4.56, 7.46, and 13.68), three outliers on page 2 (*z*s = 3.41, 6.44, and 12.69), and seven outliers on page 3 (*z*s = 5.57, 6.20, and 7.12 [x4]). There were seven unique outliers excluded for the maximum LongString index. For the page time variable, there were two outliers on page 1 (*z*s = 7.41 and 15.13), two outliers on page 2 (*z*s = 11.01 and 12.52), and four outliers on page 3 (*z*s = 4.45, 5.17, 5.28, and 14.61). There were seven unique outliers excluded for page time. See Table 4 in Appendix O for the *F*-values, partial η^2 s, means, and standard deviations for each webpage after outliers were excluded.

For the instructed response item ANOVA, sphericity was not violated. For

the Maximum LongString and Page Time, sphericity was violated based on Mauchly's test of sphericity. Therefore, the Greenhouse-Geisser corrected F -values are reported for the maximum LongString and page time analyses. There was still a nonsignificant effect for webpage on the instructed response items after excluding outliers. The significant effect for the honor code condition on the maximum LongString index remained significant, but accounted for a larger percentage of variance explained (from .033 to .083). There was a significantly greater mean maximum LongString index on page 2 and on page 3 compared to on page 1. The mean maximum LongString index was significantly greater on page 2 than was the mean maximum LongString index on page 3. The previous nonsignificant effect for page time was significant after excluding outliers. The mean page time on page 1 was significantly greater than on page 3. The mean page time on page 2 was significantly greater than on page 3. The remaining pairwise comparisons for page time were nonsignificant. It is important to note that the content of the webpages were not the same across the three webpages. This issue is addressed in Chapter 4.

The Effects of an Honor Code and Watchful Eyes on Survey Responses

Research Question #4. Eighteen one-way fixed effects ANOVAs were conducted to test Research Question #4. The ANOVAs were conducted to determine if peoples' responses to the personality measures were susceptible to heightened self-awareness (e.g., social desirability, honesty/humility). In other words, would peoples' scores be "inflated" because of the honor code or watchful

eyes? These ANOVAs served as final manipulation checks of the study's design. The same five honor code conditions were used (control condition, read-only, type-only, read-and-eyes, and type-and-eyes). Tukey's HSD tests were used to compare pairwise comparisons among the conditions for each of the personality measures.

Normality was assessed for each personality measure by calculating a z score for skewness for each honor code condition. A significantly skewed variable was defined as having a z score for skewness ± 3.3 . In the control condition, scores on the MACH-IV, DOSP-ERT-Ethical subscale, and the SRP-III were significantly positively skewed. Scores on these measures in the other four conditions and the remaining scores in all five conditions were not skewed. These positively skewed z scores indicated that most people had lower scores on these personality measures and that there were few people with higher scores on these personality measures. Because ANOVAs can be robust to violations of normality (Howell, 2010), I did not transform the data for ease of interpretation. The assumption for homogeneity of variance was violated for the extraversion scale (Levene's test = 2.45, $p = .046$), the self-esteem scale (Levene's test = 2.46, $p = .045$), and the BIDR impression management scale (Levene's test = 3.22, $p = .013$). Because Levene's test is sensitive to large samples, I also calculated the ratio of the condition with the largest variance to the condition with the smallest variance and determined if it was greater than 4:1. For the extraversion (ratio $\approx 2:1$), self-esteem (ratio $\approx 2:1$), and BIDR impression

management scale (ratio \approx 2:1), the ratio was less than 4:1. Based on this ratio, I performed the parametric one-way ANOVAs and did not make any corrections for normality or violations of homogeneity of variance. See Table 5 in Appendix O for the *F*-values, partial η^2 s, means, and standard deviations for each condition. There were no significant effects on scores for any of the 18 personality measures based on the honor code conditions.

Correlations among Careless Responding Indices, Participant Engagement, and Personality Measures

I assessed the relationships among the three careless responding indices, the two participant engagement subscales, and the personality measures. These analyses were conducted to test Research Question #5 and Research Question #6. For Research Question #5 I sought to determine what personality traits would be related to the careless responding indices. For Research Question #6, I wanted to determine how the platform (i.e., computer, tablet, cellular phone) of completing a study related to careless responding. See Table 6 in Appendix O for the correlations between each of the variables.

Correlations among Careless Responding Indices and Participant Engagement. There was a significant positive relationship between the instructed response items and the average LongString index. This correlation indicated that as the number of failed instructed response items increased, the average LongString index increased. There was a significant negative relationship between the instructed response items and the diligence scale. This correlation

indicated that as the number of failed instructed response items increased, scores on the diligence scale decreased. There was a significant negative relationship between the average LongString index and the diligence scale. This correlation indicated that as the average LongString index increased, scores on the diligence scale decreased. There was a significant positive relationship between the diligence scale and the interest scale. This correlation indicated that as scores on the diligence scale increased, scores on the interest scale increased. The UseMe item was significantly negatively correlated with the instructed response items and the average LongString index. These correlations indicated that when people reported their data should not be used, there were a greater number of instructed response items failed and a greater average LongString index. The UseMe item was significantly positively correlated with self-reported diligence and self-reported interest. These correlations indicated that when people reported their data should not be used, there were lower scores on the diligence scale and lower scores on the interest scale.

In order to compare how the platform used to complete a study related to careless responding and participant engagement, I created two dummy coded variables with computer users compared to cellular phone users and computer users compared to tablet users. There was a significant positive relationship between the computer versus cell phone dummy coded variable and instructed response items. This positive correlation indicated that people who reported completing the study on a cell phone had a higher number of failed instructed

response items than did people who reported completing the study on a computer. There was also a significant negative relationship between the computer versus cell phone dummy coded variable and the diligence scale. This negative correlation indicated that people who reported completing the study on a cell phone reported less diligence than did people who reported completing the study on a computer. The significant negative relationship between reporting technological issues and using a computer versus cellular phone is discussed in further detail in Chapter 4. There were no other significant correlations among the careless responding indices or with the participant engagement subscales.

Correlations among Careless Responding Indices and Personality

Measures. There were no significant relationships between survey time and any of the personality measures. There were a number of significant relationships among the instructed response items and the personality measures, and among the average LongString index and the personality measures. There was a significant positive relationship between the instructed response items and the following personality measures: (a) Machiavellianism, (b) DOSPERT-Financial subscale, (c) DOSPERT-Ethical subscale, (d) SDO scale, and (e) Self-Reported Psychopathy. As the number of failed instructed response items increased, scores on the above scales increased. These personality measures reflect malevolent or risk-taking personality traits. There was a significant negative relationship between the instructed responding items and the following personality measures: (a) Honesty/Humility, (b) Conscientiousness, (c)

Emotionality, (d) Self-Esteem, and (e) the BIDR Self-Deceptive Enhancement subscale. As the number of failed instructed response items increased, scores on the above scales decreased. These personality measures reflect benevolent personality traits and socially desirable traits.

I assessed the relationship between the average LongString index and the personality measures. There was a significant positive relationship between the instructed responding items and the following personality measures: (a) Machiavellianism, (b) DOSPERT-Financial subscale, (c) DOSPERT-Ethical subscale, (d) SDO scale, and (e) Self-Reported Psychopathy. As the average LongString index increased, scores on the scales listed above increased. These personality measures reflect malevolent or risk-taking personality traits. There was a significant negative relationship between the average LongString index and the following personality measures: (a) Conscientiousness, (b) DOSPERT-Social subscale, and (c) Self-Esteem. As the average LongString index increased, scores on the scales listed above decreased. Conscientiousness and self-esteem reflect positive personality traits. It is unclear, however, why the average LongString index was negatively correlated with the DOSPERT-Social subscale and is likely a Type I error. At least one Type I error is possible in this study given the large number of analyses performed and correlations computed.

Correlations among Participant Engagement and Personality Measures.

There were a number of significant relationships between the participant engagement subscales and the personality measures. I correlated the diligence

scale with the personality measures. There was a significant positive relationship between the diligence scale and the following personality measures: (a) Honesty/Humility, (b) Extraversion, (c) Conscientiousness, (d) DOSPERT-Social subscale, (e) Self-Esteem, and (f) the BIDR Self-Deceptive Enhancement subscale. As scores on the diligence scale increased, scores on the above scales increased. The personality measures reflect benevolent personality traits, with the exception of the DOSPERT-Social subscale. There was a significant negative relationship between the diligence scale and the following personality measures: (a) Machiavellianism, (b) DOSPERT-Financial subscale, (c) DOSPERT-Health/Safety subscale, (d) DOSPERT-Ethical subscale, (e) RWA scale, (f) SDO scale, and (e) Self-Reported Psychopathy. As scores on the diligence scale increased, scores on the above scales decreased. These personality measures reflect malevolent personality traits or risk-taking behaviors.

I compared the interest scale with the personality measures. There was a significant positive relationship between the interest scale and the following personality measures: (a) Honesty/Humility, (b) Extraversion, (c) Agreeableness, (d) Conscientiousness, (e) Openness, (f) Self-Esteem, (g) the BIDR Self-Deceptive Enhancement subscale, and (h) the BIDR Impression Management subscale. As scores on the interest scale increased, scores on the above scales increased. These personality measures reflect positive personality traits. There was a significant negative relationship between the interest scale and the

following personality measures: (a) Machiavellianism, (b) DOSPERT-Health/Safety subscale, (c) DOSPERT-Ethical subscale, (d) SDO scale, and (e) Self-Reported Psychopathy. As scores on the interest scale increased, scores on the above scales decreased. These personality measures reflect “dark” or malevolent personality traits.

I compared the UseMe item with the personality traits. There was a significant negative relationship with the DOSPERT-Ethical subscale and Self-Reported Psychopathy. These correlations indicated that when people reported their data should not be used in the analyses, they had higher scores on the two personality measures. These personality measures reflect malevolent and risk taking personality traits.

Method of Completing Study and Technological Issues

Participant’s answers to the items measuring the type of device used to complete the study and a yes/no question whether they experienced technological issues were used to answer Research Question #6. Additional analyses were also conducted on the UseMe variable. The goal of research question #6 was to determine how the platform (i.e., computer, tablet, cellular phone) of completing a study related to careless responding.

There were 264 people (86.6%) who reported completing the study on a laptop or desktop computer, 37 people (12.1%) who reported completing the study on a tablet, and four people (1.3%) who reported completing the study on a cellular phone. There were 12 participants (3.9%) who reported experiencing

technological issues while completing the study. Nine people provided qualitative reasons for their technological issues, including: (a) the webpage loading slowly ($n = 4$), (b) previously completed items showing as not completed or receiving a prompt stating not all items were completed ($n = 3$), laptop freezing while completing the survey ($n = 1$), and accidental user error on a tablet ($n = 1$). There were 11 people (3.6%) people who reported that their data should not be used in the analyses.

I conducted three chi square analyses comparing frequencies between the three categorical variables (the device used to complete study variable, technological issues variable, and the UseMe variable). First, I conducted a chi square analysis on the technological issues variable and device used to complete study variable. There was a nonsignificant association between the technological issues variable and device used to complete study variable, $\chi^2(2, N = 305) = 2.06, p = .357$.

I conducted a chi square analysis on the UseMe variable and device used to complete study variable. There was a significant association between the UseMe variable and device used to complete study variable, $\chi^2(2, N = 305) = 8.12, p = .017$. I then conducted a chi square analysis on the UseMe variable comparing tablet users and computer users. There was a nonsignificant association between the UseMe variable and tablets and computers users, $\chi^2(1, N = 301) = 3.01, p = .083$. I then conducted a chi square analysis on the UseMe variable comparing cellular phone and laptop/desktop computer users. There

was a significant association between the UseMe variable and cellular phone and laptop/desktop computer users, $\chi^2(1, N = 268) = 6.80, p = .009$. Out of the four people who reported using a cellular phone to complete the study, three people reported their data should be used and one person reported his or her data should not be used. Out of the 264 people who reported using a laptop/desktop computer to complete the study, 257 people reported their data should be used and seven people reported their data should not be used. The odds of people reporting their data should not be used was 12.24 times higher if they reported using a cellular phone than if they had reported using a laptop/desktop computer. This large odds ratio was likely due to the few people who reported completing the study using a cell phone. I then conducted a chi square analysis on the UseMe variable comparing cellular phone and tablet users. There was a nonsignificant association between the UseMe variable and cellular phone and tablet users, $\chi^2(1, N = 41) = 1.17, p = .279$.

I conducted a chi square analysis on the technological issues variable and the UseMe variable. There was a significant association between the technological issues variable and the UseMe variable, $\chi^2(1, N = 305) = 6.13, p = .013$. Out of the 12 people who reported experiencing technological issues, 10 people reported their data should be used and two people reported their data should not be used. Out of the 293 people who reported they did not experience technological issues, 284 people reported their data should be used and nine people reported their data should not be used. The odds of people reporting their

data should not be used was 6.31 times higher if they reported experiencing technological issues than if they had not reported experiencing technological issues.

Post Hoc Analyses to Research Question #2 and Research Question #5

The greater careless responding in the read-only condition compared to the control condition was surprising. It appeared that there might have been psychological reactance in the read-only condition (Brehm & Brehm, 1981). Additionally, many of the personality variables that would practically be suspected to correlate with the careless responding indices were correlated with the instructed response item variable when all five conditions were collapsed together (see Table 6). In the current study, the instructed response items were an overt method to detect careless responders. Participants were unaware they were being timed or that I would later be detecting how many options they selected in a consecutive row for the survey items (LongString index). Because the instructed response items were an overt method, it could also serve as a proxy for reactance in the current study. To further investigate the possibility of psychological reactance occurring, I computed the correlations between the personality measures within the control condition and the read-only condition for the instructed response items. I then conducted Fisher r -to- z transformations for independent samples. Following these transformations, I also conducted post hoc power analyses for each of the transformations using G*Power 3.1 (Faul et al., 2009). See Table 7 for the correlations for the entire sample ($N = 305$; also

provided in Table 6), the correlations for the control condition ($n = 62$), the correlations for the read-only condition ($n = 59$), the Fisher r -to- z transformation values, and observed power. The correlations for the entire sample are provided in the table for reference to the two conditions. No statistical comparisons can be calculated between either of the two conditions with the entire sample because it would violate assumptions of independence. Based on the Fisher r -to- z transformation values, there was a significant difference in the relationship between the instructed response items and emotionality between the control condition and read-only condition. In the control condition, there was a nonsignificant correlation between the instructed response items and emotionality, whereas in the read-only condition there was a significant negative relationship between the instructed response items and emotionality. Notably, observed power was low, which was likely based on the small sample sizes within each of the conditions.

CHAPTER FOUR

DISCUSSION

Introduction

Careless responders are a major concern for researchers. Participants engaging in this behavior can cost a researcher time and money and lead to a researcher having unpublishable data. Although it is important for researchers to detect and eliminate these cases, it is also important for researchers to determine ways to reduce this behavior before it occurs. The purpose of the present study was to examine the effectiveness of increasing people's self-awareness with the use of an honor code and the presentation of a picture of eyes before completing a survey in order to reduce careless responding. I tested two hypotheses and six research questions. I will discuss the conclusions based on the data and directions for future research, limitations of the study's design and methodology, and a brief conclusion of the study.

Conclusions to Hypotheses and Research Questions and Directions for Future Research

First, I will interpret Hypothesis 1, Research Question #1, and Research Question #2 together because of the overlap with the analyses conducted. For Hypothesis 1, I predicted that careless responding would be reduced in the honor code conditions that involved typing an honor code and/or in conjunction with viewing a picture of a pair of eyes (type-only, read-and-eyes, and type-and-eyes)

compared to the control condition or the read-only condition. Specifically, I predicted that there would be significantly fewer instructed response items failed in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition (H1a). I predicted that there would be a smaller average LongString index in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition (H1b). I predicted that there would be more time spent completing the survey in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition (H1c). For Research Question #1, I investigated if one of the three primary conditions (type-only, read-and-eyes, and type-and-eyes) that I predicted to be the most effective at reducing careless responding would be more effective than one of the other two primary conditions at reducing careless responding. For Research Question #2, I investigated whether careless responding would be significantly reduced in the read-only condition compared to the control condition.

The results did not support my predictions for Hypotheses 1a, 1b, and 1c. The three conditions (type-only, read-and-eyes, and type-and-eyes) did not reduce careless responding in comparison to the control condition or the read-only condition on the three careless responding indices. Because there was not a reduction in careless responding in the three primary conditions, it was unnecessary to further investigate the differences among conditions (i.e.,

Research Question #1).

Because of the lack of support for Hypothesis 1, I compared the results of my study to Meade and Craig's (2012, p. 443, Table 3) results. In their study, they used 10 bogus items, whereas I used 12 instructed response items in my study. The mean score for the instructed response items in my control condition had a similar mean score for the bogus items in Meade and Craig's study in the identified condition (1.00 in my study vs. 0.93 in their study). Their control condition ($M = 1.50$) was higher than what I found in my control condition. Maniaci and Rogge (2014, Study 1) found a mean score of 0.32 when using seven instructed response-like items in a large sample of undergraduate students. Because these rates vary across samples, it is possible that there might be situational factors or institutional policies that impact rates of careless responding. These factors and policies could influence participants' willingness or desire to engage in careless responding. The policies might include other methods available to students to obtain extra credit, possibly losing credit for carelessly responding in research, among others. The social situation of students (e.g., SES of the students, children in the home) might also vary from campus to campus, which might also play a role in engaging in careless responding. These factors might be important to investigate and consider in future research.

It was surprising to find an increase in careless responding in the read-only condition compared to in the control condition on the instructed response item variable and average LongString index. There was also a marginally

significant difference between the read-only condition and the read-and-eyes condition, and the read-only condition and the type-and-eyes condition. In both cases there was a greater number of instructed response items failed in the read-only condition. This increase in careless responding possibly reflects psychological reactance from participants in the read-only condition. This is a plausible explanation because the instructed response items were an overt method to detect careless responders. Brehm and Brehm (1981) defined psychological reactance as when "...a threat to or loss of a freedom motivates the individual to restore that freedom" (p. 4). A reactance-like response is likely to occur to any behavior that is perceived to be restricted or threatened by someone else, in this case the researcher. People will lash out or react negatively to restore perceived loss of power or control over a situation. In my research, the participants might have experienced reactance if they felt the researcher was trying to control their responses or implied they were dishonest.

The potential reactance effect should be prevalent among some individuals more so than others. After additional post hoc analyses (see Table 7), I found a number of interesting differences between the instructed response items and personality measure correlations between the control condition and read-only condition. Although only one of the Fisher *r*-to-*z* transformations was significant, there were also some interesting trends. First, it is unclear why the emotionality scale was different between the two conditions based on items' content. These items do not relate to reactance either theoretically or on face

value. Some of the other measures, however, such as self-reported psychopathy, Machiavellianism, self-esteem, RWA, the DOSPERT-Ethical, BIDR impression management, and conscientiousness might be reflective of a reactance-like behavior observed in the study. My only finding one significant *r*-to-*z* transformation might be due partly to the small sample size and low observed statistical power. Future research could seek to replicate these findings and further explore this possible reactance-like behavior with a proper sample size.

The honor code in the current study was written in a way to hopefully induce a sense of personal moral identity and self-awareness (e.g., “I agree to read all instructions...”). The way in which the honor code was written, however, might have backfired and been what led to the reactance-like effect in the read-only condition. This finding might be important for researchers who are considering including an honor code statement in their future surveys. If researchers use an honor code in a completely anonymous situation, which was the case in the current study, they might negatively impact their results inadvertently. Participants might respond more carelessly than if the researchers did not include an honor code. Future studies could adjust the honor code to focus on the researcher rather than on the participant (e.g., “We, the researchers, ask that all of the instructions and survey items be read carefully to help us with our study”). This change in wording might remove this reactance-like effect. Future research could also examine the effects of different types of honor codes in anonymous and nonanonymous conditions.

Considering there was a reactance-like increase in careless responding in the read-only condition, it was also surprising there was not a similar increase in the read-and-eyes condition. One possible reason is that the picture of eyes neutralized the negative effects of the honor code. This explanation, however, cannot be concluded from the present study alone. There might still be an important mechanism to the presence of eyes with survey responses. Because eyes can increase prosocial behavior in economic games (Haley & Fessler, 2005) and field settings (Bateson et al., 2006; Ernest-Jones et al., 2011), it would be important to investigate the effects of only a picture of a pair of eyes, without an honor code, on response behavior as well. Ernest-Jones et al. (2011) found that people were more likely to pick up their trash in a cafeteria with a poster of a picture of eyes. This increase in trash thrown away was present regardless of whether a poster was presented with a relevant message explaining that someone should pick up his or her trash or when the poster was presented with an irrelevant message. This finding implies that the presence of a picture of eyes might have an effect on behavior without a relevant instructional message presented with the picture. Although the current study was primarily focused on the use of an honor code in conjunction with a picture of a pair of watchful eyes, the results provide some preliminary evidence that the presence of watchful eyes could still hold some effects on careless responding behavior. Recall, that the eyes (in the read-and-eyes condition) appeared to have neutralized the negative effects found in the read-only condition.

The presence of a pair of watchful eyes could also remedy one of the limitations associated with online research. Meade and Craig (2012) noted that one concern with online research is the loss of social contact between the researcher and participant. A picture of watchful eyes or a researcher at the start of a survey, or on each webpage of a survey, might “bring the researcher to the respondent” and increase effortful responding. Research on the identifiable victim effect supports this prediction. For example, Jenni and Loewenstein (1997) found that participants were more likely to help a person in need if they viewed a picture of that person rather than just reading about that person’s plight. This finding in the context of the current study indicates that a picture of the researcher could encourage participants to see the researcher as a person in need of their help. This “identifiable researcher effect” could motivate participants to put forth additional effort when completing a survey thereby reducing careless responding.

For the analyses involving the survey time variable, there were no significant differences across conditions and the variable was highly skewed. The variable also did not correlate with the other two careless responding indices, the participant engagement subscales, the single item UseMe variable, or the personality measures. Some researchers (see Huang et al., 2012) have stated that this index can be an effective method for detecting careless responders by developing certain cut points. Maniaci and Rogge (2014), however, found that removing careless responders based on survey time measures did not result in a

large increase in statistical power, whereas other methods to remove careless responders (i.e., instructed response items) improved statistical power to a greater extent. Yan and Tourangeau (2008) also found that factors such as education and experience with the Internet have an impact on the amount of time it takes a respondent to complete a survey. In their study, participants who reported having more education and more experience with the Internet completed surveys faster than did people who reported less education and less experience with the Internet. Therefore, survey time might not be the best measure to identify careless responders.

Response times to single survey items might be a better behavior to investigate in future research. One method to determine these response times could be with the use of eye tracking equipment. Galesic, Tourangeau, Couper, and Conrad (2008) conducted three experiments investigating survey response behavior with eye tracking equipment. They found that people generally engage in short cuts such as reading the first few survey items carefully and then reading the remaining items quickly, or not reading response options carefully when there are a lot of options available to choose from. Understanding how people read survey items might help to understand other factors that might be related to survey responses and misinterpretations of survey items.

For Hypothesis 2, I predicted that there would be significantly higher scores on the diligence scale and the interest scale in the honor code conditions that involved typing an honor code and/or in conjunction with viewing a picture of

a pair of eyes (type-only, read-and-eyes, and type-and-eyes) compared to the control condition or the read-only condition. Specifically, I predicted that there would be significantly higher scores on the diligence scale in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition (H2a). I predicted that there would be significantly higher scores on the interest scale in the type-only condition, read-and-eyes condition, and type-and-eyes condition compared to the control condition or the read-only condition (H2b).

The results did not support my predictions for Hypotheses 2a and 2b. People in the three conditions (type-only, read-and-eyes, and type-and-eyes) did not report significantly more diligence or interest completing the survey. For the diligence scale, I found a ceiling effect on this measure. This ceiling effect was reflective of the negative skewness in four of the five honor code conditions and by observing the mean scores in each condition. This ceiling effect was likely the reason there was a nonsignificant difference on this variable. There was a marginally significant difference for the interest scale in this study. Meade and Craig (2012) found a marginally significant difference for self-reported interest across their three conditions. When people were given a stern warning to complete the survey carefully and honestly and typed in their name, they reported greater interest. The results in the current study were trending in this direction, but to less of a degree than what was observed in Meade and Craig's study.

For Research Question #3, I investigated if careless responding increased in the middle and at the end of a survey or if careless responding remained consistent throughout the survey. Because the analyses with outliers removed was more informative, I will focus primarily on those analyses. For Research Question #3, there appeared to be little to no difference on the careless responding indices across the webpages. To address this question, the best careless responding index was the instructed response items, which was not significantly different across the three webpages. The maximum LongString index and survey time can be easily influenced by the types of scales displayed on each page. For example, the types of scale might influence responses and higher maximum LongString index scores might reflect differences with specific scales, especially if consecutive items are all positively worded or all negatively worded. In the current study, the greatest maximum LongString index was observed on page 2. The difference in the maximum LongString index was likely reflective of the DOSPERT scale items all being worded in the same direction.

Differences in survey content were also likely the explanation for the differences observed in time spent on each of the three webpages. Heerwegh (2003; Yan & Tourangeau, 2008) reported that attitude strength can increase the amount of time a person takes to respond to a survey item that assesses attitudes. Because the second page contained a number of attitudinal scales, such as the RWA and SDO scales, people might have taken longer to complete these scales.

The dropout rates across the three webpages might also be relevant to Research Question #3. Participant dropout rates increased slightly across the three webpages. Although no strong conclusions can be made based on these results, the dropout rates might reflect a similar trend for experiencing fatigue, a loss of diligence, or a loss of interest while completing long surveys. Future studies could examine how participant characteristics and careless responding relate to participant retention.

Based on the results of the current study, there did not appear to be greater careless responding across the survey. This finding, however, should be further examined to see at what point in long surveys people might be likely to engage in careless responding. Berry et al. (1992) found that people self-reported engaging in careless responding in the middle to end of the survey. In their study, participants completed the MMPI-2, which includes close to 600 true/false survey items, whereas in the current study people completed roughly 300 survey items. Future studies could increase the number of items and webpages to determine at what point in a study careless responding begins to increase. This information could help researchers determine limits to survey length for obtaining higher quality data from participants.

For Research Question #4, I investigated whether the conditions predicted to reduce careless responding (type-only condition, read-and-eyes condition, type-and-eyes condition) would lead people to “inflate” their scores to certain personality measures that may be susceptible to heightened self-awareness

(e.g., social desirability, honesty/humility). Because nearly all of the personality measures in the study could be influenced by heightened self-awareness, I examined differences for all 18 personality measures. In the current study, there were no differences in scores on the personality measures across the honor code conditions. This finding, although encouraging, might simply be due to people not being inclined to respond in a socially desirable manner because the honor code and watchful eyes did not have an effect on self-awareness. Interestingly, the DOSPERT Social subscale was marginally significant. There were higher scores in the control condition compared to the other conditions. Based on the number of tests, it is possible that there might be some false positives (i.e., Type I error). Replicating the study would be necessary to determine if this finding was a Type I error or a concern requiring further investigation.

For Research Question #5, I examined if the careless responding indices would be correlated with scores on the personality measures. Scores on many of the personality measures were correlated with the instructed response items, average LongString index, and participant engagement subscales. Many of these personality measures provide initial construct validity for instructed response items, the average LongString index, and the participant engagement scales. Because many of the personality measures are reflective of both benevolent traits (e.g., honesty, conscientiousness) and malevolent traits (e.g., Machiavellianism, ethical risk-taking), these relationships might reflect types of

people who are detected by the careless responding indices. Specifically, these measures are reflective of people who pay close attention to details (conscientiousness), are honest/dishonest (honesty/humility and DOSPERT-Ethical subscale), and not trying to take advantage of the situation (Machiavellianism and psychopathy). Survey time was unrelated to the other careless responding indices, participants' scores on the engagement scales, and the personality measures. As mentioned above, response time might be reflective of factors other than careless responding such as reading speeds, education level, or familiarity with using the Internet.

For Research Question #6, I examined how the platform (i.e., computer, tablet, cellular phone) of completing a study was related to careless responding. In the current study, roughly 13 percent of the sample reported completing the survey on a device (cellular phone or tablet) other than a desktop or laptop computer. For tablet users compared to computer users, there appeared to be no significant relationships with the careless responding indices or reporting experiencing technological issues. For cellular phone users compared to computer users, however, there was a significant positive relationship with higher careless responding for the instructed response items and the UseMe variable. It is important to consider that the number of people who reported completing the study on a cellular phone was small; the results from these analyses are likely more reflective of the small sample size in these analyses.

Screen size might also be a factor for the significant relationship between

the dummy coded cellular phone versus computer variable and careless responding. Because the current study did not utilize a responsive web design, the webpage and survey likely did not adjust properly to tablet and cellular phone screens. Tablet users might not experience the same issues as someone with a cellular phone because of the larger screen size. The smaller screen sizes with cellular phones are likely difficult to use for properly reading, understanding, and responding to survey items. This difficulty in completing the study might reflect misclicking another option on the instructed response items or participants putting forth less effort. The explanation of less effort might be indicative of the significant negative relationship between the dummy coded variable and the diligence scale. This preliminary evidence is greatly limited by the small portion of the sample completing the study with a cellular phone. As cellular phone programmers are developing more capabilities for phones and people are beginning to use cellular phones for more activities, it might be important to track their use for completing online studies. If their use for completing studies increases, researchers might seek to investigate ways to increase data quality for people using these devices. For example, researchers could randomly assign participants to complete a study on either a computer or cellular phone. Response behavior could then be assessed for measurement errors associated with the device used to complete the study.

Limitations of Study Design and Procedures

In the current study, there were a number of limitations that should be considered when interpreting the results. First, the instructed response item variable might not be a meaningful proxy for psychological reactance. Without providing sufficient validation, this possibility limits the generalizability of the reactance interpretation. With the careless responding measures that I used in the study, the instructed response item measure was the most practical to use because its purpose to detect careless responding was likely clear to participants. The average LongString index was also greater in the read-only condition providing additional support for a reactance-like behavior in the read-only condition. Future research should examine reactance under similar conditions through additional means such as using self-report items at the end of the survey.

Second, the use of a between-person design for determining the effects of the honor code on responses to the personality scales requires further testing with additional research designs. People who are more susceptible to honor codes and watchful eyes might have responded in a self-inflating or socially desirable manner because of the heightened self-awareness. Future studies could benefit by using a repeated measures design with a one- to two-week break between completing the personality surveys to investigate this possible socially desirable responding further. By including additional items for detecting peoples' susceptibility to honor codes and watchful eyes, researchers can identify these cases and discern whether they might threaten the reliability of the

data. Additionally, if people were randomly assigned to read an honor code at either time 1 or time 2, and not the other time they completed the study, it would allow researchers to determine if people engaged in a reactance-like behavior.

Third, correlating the personality measures with the careless responding indices might not allow for meaningful interpretations. Berry et al. (1992), across four studies, found that half or more of their samples were carelessly responding to one or a few items while completing a survey. However, none of the sample in one study and up to seven percent in another study reported responding carelessly to most or all of the items. It is likely that there were similar rates in the current study. This is also reflective of the small number of people who self-reported that their data should not be used in the analyses. Furthermore, personality measures that would be expected to relate to the careless responding indices were in fact related in directions that would be meaningful. Future studies can seek to replicate these findings to determine if the significant relationships have a meaningful interpretation or were simply due to chance.

Last, the sample size was small to make strong inferences for Research Question #6 and the post hoc analyses investigating the possibility of psychological reactance. The power analysis I conducted prior to the study was for investigating my two primary hypotheses, which was the focus of this study. Beyond those two initial predictions, the remaining research questions were exploratory and merely as a guide for future research. The results based on the research questions might provide directions for meaningful investigations.

Conclusion

The current study tested different methods of using an honor code along with a picture of a pair of watchful eyes to reduce careless responding. The findings indicate that an honor code might have negative effects on response behavior and lead to psychological reactance. Watchful eyes, however, is one method that warrants future research. In the current study, the relationships between personality and careless responding measures provide some initial evidence of construct validity for measures to detect careless responding.

The results from the current study also hold theoretical and practical implications. Under certain circumstances, honor codes might lead to negative outcomes. Previous research has reported on the benefits of using honor codes (McCabe et al., 2002; McCabe & Trevino, 1997; Shu et al., 2011), whereas the current study found a possible circumstance where the use of an honor code could backfire. These situational differences should be further examined to determine under what circumstances honor codes lead to heightened self-awareness and positive outcomes and when honor codes lead to psychological reactance and negative outcomes. Factors such as signing one's name, anonymity of the situation, phrasing of the honor code, and incentives/repercussions for engaging in the behavior might be factors researchers could further explore. These future research directions would be meaningful for creating online surveys as well as important for educational

institutions and organizations.

Additionally, it is important for researchers to explore what further methods would be optimal for reducing careless responding and their implementation. Methods to reduce careless responding that show promise at one institution in future investigations, should also be examined for effectiveness at other institutions. If differences in institutional policies or norms relate to people engaging in careless responding, certain methods might be ineffective at one institution, but not another. In sum, researchers could benefit more greatly with a focus on finding methods to reduce careless responding rather than on methods to detect and removed careless responders.

APPENDIX A
INFORMED CONSENT FORM

INFORMED CONSENT FORM



College of Social and Behavioral Sciences
Department of Psychology

Informed Consent Personality Assessment

PURPOSE: The study in which you are being asked to participate requires you to complete a number of personality measures. This study is being conducted by Alexander S. McKay, a graduate student at California State University, San Bernardino (CSUSB), and Donna M. Garcia, Associate Professor of Psychology at CSUSB. This study has been approved by the Department of Psychology Institutional Review Board Subcommittee of the California State University, San Bernardino, and this consent form should bear the official Psychology IRB stamp of approval. The University requires that you give your consent before participating in this study.

DESCRIPTION: If you choose to participate in the study, you will be asked to complete a number of personality tests. The survey will take around 1 hour and 30 minutes. You will be compensated with 3 extra credit points.

PARTICIPATION: Participation in this study is voluntary. You may choose to participate or not. If you choose to participate but later change your mind, you may withdraw from the study at any time. Refusal to participate or withdrawal at any time during the study will involve no penalty or loss of benefits to which you are otherwise entitled. You are not waiving any legal claims, rights, or remedies because of your participation in this research study.

CONFIDENTIALITY OR ANONYMITY: Details about your performance on the tasks and your responses on the survey will be used solely by the researchers and stored on a secure computer or locked in laboratory cabinet, with no identifying information about you attached. By signing this form you give permission for the use of your data to be published in aggregate form by the researcher. Data will be destroyed five years after publication. You have the right to cancel your permission to use and disclose information pertaining to your participation in the study, in writing, at any time, by sending your written request to Dr. Donna Garcia (dmgarcia@csusb.edu).

DURATION: Your participation in the study will take approximately 1 hour and 30 minutes.

RISKS: There are no known risks to participating in this study. These tests should cause no more discomfort than you would experience in your everyday life.

BENEFITS: Although participation may not benefit you directly, we believe that the information obtained from this study will help us gain a better understanding of how individuals respond in evaluative situations.

QUESTIONS: If you have questions about the research or your rights as a research subject, or if you wish to learn about the results of this study, please contact Dr. Donna Garcia at 909-537-3893 or dmgarcia@csusb.edu. This study has been approved by the Institutional Review Board at California State University, San Bernardino, and a copy of the official California State University, San Bernardino Institutional Review Board stamp of approval should appear on this consent form.

I acknowledge that I have been informed of, and that I understand the nature and purpose of this study, and I freely consent to participate. I also acknowledge that I am at least 18 years of age.

California State University			
Psychology Institutional Review Board Sub-Committee			
Approved	2/28/14	Void After	2/28/15
IBB #		Chair	
H14WI-24			

The California State University
Bakersfield • Channel Islands • Chico • Dominguez Hills • East Bay • Fresno • Fullerton • Humboldt • Long Beach • Los Angeles • Maritime Academy
Monterey Bay • Northridge • Pomona • Sacramento • San Bernardino • San Diego • San Francisco • San Jose • San Luis Obispo • San Marcos • Sonoma • Stanislaus

APPENDIX B
HONOR CODE AND WATCHFUL EYES

HONOR CODE AND WATCHFUL EYES

Instructions on Honor Code Page:

For this study, you will complete eight different personality measures broken up onto three webpages.

Honor code:

Completing this study honestly is part of CSUSB's academic ethics code. Therefore, I agree to read all instructions and respond to all of the survey items carefully. I will not provide responses to survey items that do not reflect my true behavior. Providing good quality responses is important to the field of psychology and I wish to do my part by answering truthfully.

Watchful Eyes:

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2, 412-414.

doi:10.1098/rsbl.2006.0509



APPENDIX C
CARELESS RESPONDING ITEMS

CARELESS RESPONDING ITEMS

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455. doi:10.1037/a0028085

Additional careless responding assessment items:

1. Select strongly agree to this item.
2. Select strongly disagree to this item.

The underlined portion of these items will vary throughout the survey. They may read: “agree,” “disagree,” “strongly disagree,” etc.

APPENDIX D
PARTICIPANT ENGAGEMENT MEASURES

PARTICIPANT ENGAGEMENT MEASURES

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455. doi:10.1037/a0028085

The diligence scale is composed of items 1-9 and the interest scale is composed of items 10-15. Items with an (R) were reverse-coded items.

Please complete the items using the following scale:

1 = strongly disagree

7 = strongly agree

Participant Engagement Scale

1. I carefully read every survey item.
 2. I could've paid closer attention to the items than I did. (R)
 3. I probably should have been more careful during this survey. (R)
 4. I worked to the best of my abilities in this study.
 5. I put forth my best effort in responding to this survey.
 6. I didn't give this survey the time it deserved. (R)
 7. I was dishonest on some items. (R)
 8. I was actively involved in this study.
 9. I rushed through this survey. (R)
 10. I enjoyed participating in this study.
 11. This study was a good use of my time.
 12. I was bored during the study. (R)
 13. This survey was too long. (R)
 14. The work I did for this study is important to me.
 15. I care about my performance in this study.
-
1. In your honest opinion, should we use your data in our analyses in this study?
 - "yes"
 - "no"

APPENDIX E
HEXACO-100

HEXACO-100

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39*, 329-358.
doi:10.1207/s15327906mbr3902_8

The Openness scale is composed of items 1-16. The Conscientiousness scale is composed of items 17-32. The Agreeableness scale is composed of items 33-48. The Extraversion scale is composed of items 49-64. The Emotionality scale is composed of items 65-80. The Honesty/Humility scale is composed of items 81-96. The interstitial Altruism scale is composed of items 97-100. Items with an (R) were reverse-coded items.

Please complete the items using the scale provided:

1 = strongly disagree

7 = strongly agree

1. I would be quite bored by a visit to an art gallery. (R)
2. I'm interested in learning about the history and politics of other countries.
3. I would like a job that requires following a routine rather than being creative. (R)
4. I think that paying attention to radical ideas is a waste of time. (R)
5. I wouldn't spend my time reading a book of poetry. (R)
6. I enjoy looking at maps of different places.
7. I would enjoy creating a work of art, such as a novel, a song, or a painting.
8. I like people who have unconventional views.
9. If I had the opportunity, I would like to attend a classical music concert.
10. I would be very bored by a book about the history of science and technology. (R)
11. People have often told me that I have a good imagination.
12. I think of myself as a somewhat eccentric person.
13. Sometimes I like to just watch the wind as it blows through the trees.
14. I've never really enjoyed looking through an encyclopedia. (R)
15. I don't think of myself as the artistic or creative type. (R)
16. I find it boring to discuss philosophy. (R)
17. I clean my office or home quite frequently.
18. When working, I often set ambitious goals for myself.
19. I often check my work over repeatedly to find any mistakes.
20. I make decisions based on the feeling of the moment rather than on careful thought. (R)
21. I plan ahead and organize things, to avoid scrambling at the last minute.
22. I often push myself very hard when trying to achieve a goal.

23. When working on something, I don't pay much attention to small details. (R)
24. I make a lot of mistakes because I don't think before I act. (R)
25. People often joke with me about the messiness of my room or desk. (R)
26. Often when I set a goal, I end up quitting without having reached it. (R)
27. I always try to be accurate in my work, even at the expense of time.
28. I don't allow my impulses to govern my behavior.
29. When working, I sometimes have difficulties due to being disorganized. (R)
30. I do only the minimum amount of work needed to get by. (R)
31. People often call me a perfectionist.
32. I prefer to do whatever comes to mind, rather than stick to a plan. (R)
33. I rarely hold a grudge, even against people who have badly wronged me.
34. People sometimes tell me that I am too critical of others. (R)
35. People sometimes tell me that I'm too stubborn. (R)
36. People think of me as someone who has a quick temper. (R)
37. My attitude toward people who have treated me badly is "forgive and forget".
38. I generally accept people's faults without complaining about them.
39. I am usually quite flexible in my opinions when people disagree with me.
40. I rarely feel anger, even when people treat me quite badly.
41. If someone has cheated me once, I will always feel suspicious of that person. (R)
42. I tend to be lenient in judging other people.
43. When people tell me that I'm wrong, my first reaction is to argue with them. (R)
44. Most people tend to get angry more quickly than I do.
45. I find it hard to fully forgive someone who has done something mean to me. (R)
46. Even when people make a lot of mistakes, I rarely say anything negative.
47. I find it hard to compromise with people when I really think I'm right. (R)
48. I find it hard to keep my temper when people insult me. (R)
49. I feel reasonably satisfied with myself overall.
50. I rarely express my opinions in group meetings. (R)
51. I avoid making "small talk" with people. (R)
52. I am energetic nearly all the time.
53. I think that most people like some aspects of my personality.
54. In social situations, I'm usually the one who makes the first move.
55. I enjoy having lots of people around to talk with.
56. On most days, I feel cheerful and optimistic.
57. I feel that I am an unpopular person. (R)
58. When I'm in a group of people, I'm often the one who speaks on behalf of the group.
59. I prefer jobs that involve active social interaction to those that involve

- working alone.
60. People often tell me that I should try to cheer up. (R)
 61. I sometimes feel that I am a worthless person. (R)
 62. I tend to feel quite self-conscious when speaking in front of a group of people. (R)
 63. The first thing that I always do in a new place is to make friends.
 64. Most people are more upbeat and dynamic than I generally am. (R)
 65. I would feel afraid if I had to travel in bad weather conditions.
 66. I sometimes can't help worrying about little things.
 67. When I suffer from a painful experience, I need someone to make me feel comfortable.
 68. I feel like crying when I see other people crying.
 69. I don't mind doing jobs that involve dangerous work. (R)
 70. I worry a lot less than most people do. (R)
 71. I can handle difficult situations without needing emotional support from anyone else. (R)
 72. When someone I know well is unhappy, I can almost feel that person's pain myself.
 73. When it comes to physical danger, I am very fearful.
 74. I rarely, if ever, have trouble sleeping due to stress or anxiety. (R)
 75. Whenever I feel worried about something, I want to share my concern with another person.
 76. I feel strong emotions when someone close to me is going away for a long time.
 77. Even in an emergency I wouldn't feel like panicking. (R)
 78. I get very anxious when waiting to hear about an important decision.
 79. I rarely discuss my problems with other people. (R)
 80. I remain unemotional even in situations where most people get very sentimental. (R)
 81. If I want something from a person I dislike, I will act very nicely toward that person in order to get it. (R)
 82. If I knew that I could never get caught, I would be willing to steal a million dollars. (R)
 83. Having a lot of money is not especially important to me.
 84. I am an ordinary person who is no better than others.
 85. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.
 86. I would be tempted to buy stolen property if I were financially tight. (R)
 87. I would like to live in a very expensive, high-class neighborhood.
 88. I wouldn't want people to treat me as though I were superior to them.
 89. If I want something from someone, I will laugh at that person's worst jokes. (R)
 90. I would never accept a bribe, even if it were very large.
 91. I would like to be seen driving around in a very expensive car. (R)

92. I think that I am entitled to more respect than the average person is. (R)
93. I wouldn't pretend to like someone just to get that person to do favors for me.
94. I'd be tempted to use counterfeit money, if I were sure I could get away with it. (R)
95. I would get a lot of pleasure from owning expensive luxury goods. (R)
96. I want people to know that I am an important person of high status. (R)
97. I have sympathy for people who are less fortunate than I am.
98. I try to give generously to those in need.
99. It wouldn't bother me to harm someone I didn't like.
100. People see me as a hard-hearted person.

APPENDIX F
MACHIAVELLIANISM (MACH-IV)

MACHIAVELLIANISM (MACH-IV)

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York, NY:

Academic Press.

Items with an (R) were reverse-coded items.

Please complete the items using the scale provide:

1 = strongly disagree

7 = strongly agree

1. Anyone who completely trusts anyone else is asking for trouble.
2. Most people who get ahead in the world lead clean moral lives. (R)
3. It is safest to assume that all people have a vicious streak and it will come out when they are given a chance.
4. One should take action only when sure it is morally right. (R)
5. Generally speaking, people won't work hard unless they're forced to do so.
6. It is wise to flatter important people.
7. It is hard to get ahead without cutting corners here and there.
8. People suffering from incurable diseases should have the choice of being put painlessly to death.
9. Most people are brave. (R)
10. The best way to handle people is to tell them what they want to hear.
11. The biggest difference between most criminals and other people is that criminals are stupid enough to get caught.
12. Honesty is the best policy in all cases. (R)
13. Barnum was very wrong when he said there's a sucker born every minute. (R)
14. Most people are basically good and kind. (R)
15. When you ask someone to do something for you, it is best to give the real reasons for wanting it rather than giving reasons which might carry more weight. (R)
16. It is possible to be good in all respects. (R)
17. Most people forget more easily the death of a parent than the loss of their property.
18. Never tell anyone the real reason you did something unless it is useful to do so.
19. There is no excuse for lying to someone else. (R)
20. All in all, it is better to be humble and honest than to be important and dishonest. (R)

APPENDIX G
DOMAIN-SPECIFIC RISK-TAKING (DOSPERT)

DOMAIN-SPECIFIC RISK-TAKING (DOSPERT)

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15, 263-290. doi:10.1002/bdm.414

For each of the following statements, please indicate the likelihood that you would engage in the described activity or behavior, if you were to find yourself in that situation.

1 = very unlikely

7 = very likely

1. Admitting that your tastes are different from those of a friend.
2. Going camping in the wilderness.
3. Betting a day's income at the horse races.
4. Swimming far out from shore on an unguarded lake or ocean.
5. Investing 10% of your annual income in a moderate growth mutual fund.
6. Drinking heavily at a social function.
7. Taking some questionable deductions on your income tax return.
8. Disagreeing with an authority figure on a major issue.
9. Betting a day's income at a high-stake poker game.
10. Having an affair with a married man/woman.
11. Passing off somebody else's work as your own.
12. Going on vacation to a third-world country.
13. Arguing with a friend who has a different opinion on an issue.
14. Going down a ski run that is beyond your ability.
15. Investing 5% of your annual income in a very speculative stock.
16. Approaching your boss for a raise.
17. Going whitewater rafting at high water in the spring.
18. Betting a day's income on the outcome of a sporting event (e.g., baseball, soccer, or football).
19. Investing 5% of your annual income in a dependable and conservative stock.
20. Engaging in unprotected sex.
21. Revealing a friend's secret to someone else.
22. Driving a car without wearing a seat belt.
23. Investing 10% of your annual income in a new business venture.
24. Taking a weekend sky diving class.
25. Riding a motorcycle without a helmet.
26. Gambling a week's income at a casino.
27. Choosing a career that you truly enjoy over a more prestigious one.
28. Downloading proprietary software from the Internet.
29. Reporting a neighbor or friend for some illegal activity.

30. Speaking your mind about an unpopular issue in a meeting at work.
31. Sunbathing without sunscreen.
32. Bungee-jumping off a tall bridge.
33. Piloting a small plane.
34. Walking home alone at night in an unsafe area of town.
35. Eating high cholesterol foods.
36. Driving while taking medication that may make you drowsy
37. Moving to a city far away from your extended family.
38. Starting a new career in your mid-thirties.
39. Leaving your young children alone at home while running an errand.
40. Not returning a wallet you found that contains \$200.

APPENDIX H
RIGHT WING AUTHORITARIANISM (RWA)

RIGHT WING AUTHORITARIANISM (RWA)

Zakrisson, I. (2005). Construction of a short version of the Right-Wing

Authoritarianism (RWA) scale. *Personality and Individual Differences*, 39,

863-872. doi:10.1016/j.paid.2005.02.026

Items with an (R) were reverse-coded items.

Please rate the degree to which you agree with the following statements.

1 = strongly disagree

7 = strongly agree

1. Our country needs a powerful leader, in order to destroy the radical and immoral currents prevailing in society today.
2. Our country needs free thinkers, who will have the courage to stand up against traditional ways, even if this upsets many people. (R)
3. The “old-fashioned ways” and “old-fashioned values” still show the best way to live.
4. Our society would be better off if we showed tolerance and understanding for untraditional values and opinions. (R)
5. God’s laws about abortion, pornography and marriage must be strictly followed before it is too late, violations must be punished.
6. The society needs to show openness towards people thinking differently, rather than a strong leader, the world is not particularly evil or dangerous. (R)
7. It would be best if newspapers were censored so that people would not be able to get hold of destructive and disgusting material.
8. Many good people challenge the state, criticize the church and ignore “the normal way of living”.
9. Our forefathers ought to be honored more for the way they have built our society, at the same time we ought to put an end to those forces destroying it. (R)
10. People ought to put less attention to the Bible and religion, instead they ought to develop their own moral standards.
11. There are many radical, immoral people trying to ruin things; the society ought to stop them. (R)
12. It is better to accept bad literature than to censor it.
13. Facts show that we have to be harder against crime and sexual immorality, in order to uphold law and order. (R)
14. The situation in the society of today would be improved if troublemakers were treated with reason and humanity.
15. If the society so wants, it is the duty of every true citizen to help eliminate

the evil that poisons our country from within. (R)

APPENDIX I
SOCIAL DOMINANCE ORIENTATION (SDO)

SOCIAL DOMINANCE ORIENTATION (SDO)

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social Dominance Orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67, 741-763.

doi:10.1037/0022-3514.67.4.741

Items with an (R) were reverse-coded items.

Please rate the degree to which you agree with the following statements.

1 = strongly disagree

7 = strongly agree

1. Some groups of people are simply inferior to other groups.
2. In getting what you want, it is sometimes necessary to use force against other groups.
3. It's OK if some groups have more of a chance in life than others.
4. To get ahead in life, it is sometimes necessary to step on other groups.
5. If certain groups stayed in their place, we would have fewer problems.
6. It's probably a good thing that certain groups are at the top and other groups are at the bottom.
7. Inferior groups should stay in their place.
8. Sometimes other groups must be kept in their place.
9. It would be good if groups could be equal. (R)
10. Group equality should be our ideal. (R)
11. All groups should be given an equal chance in life. (R)
12. We should do what we can to equalize conditions for different groups.
13. We should do what we can to increase social equality. (R)
14. We would have fewer problems if we treated people more equally. (R)
15. We should strive to make incomes as equal as possible. (R)
16. No one group should dominate in society. (R)

APPENDIX J
SELF-ESTEEM

SELF-ESTEEM

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ:

Princeton University Press.

Items with an (R) were reverse-coded items.

Please indicate your agreement with the following statements with the scale provided.

1 = strongly disagree

7 = strongly agree

1. On the whole, I am satisfied with myself.
2. At times, I think I am no good at all. (R)
3. I feel that I have a number of good qualities.
4. I am able to do things as well as most other people.
5. I feel I do not have much to be proud of. (R)
6. I certainly feel useless at times. (R)
7. I feel that I am a person of worth, at least on an equal plane with others.
8. I wish I could have more respect for myself. (R)
9. All in all, I am inclined to feel that I am a failure. (R)
10. I take a positive attitude toward myself.

APPENDIX K

BALANCED INVENTORY OF DESIRABLE RESPONDING (BIDR)

BALANCED INVENTORY OF DESIRABLE RESPONDING (BIDR)

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P.

Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). New York, NY:

Academic Press.

The Self-Deceptive Enhancement scale is composed of items 1-20 and the Impression Management scale is composed of items 21-41. Items with an (R) were reverse-coded items.

Please rate the items using the scale provide

1 = very untrue of me

7 = very true of me

1. My first impressions of people usually turn out to be right.
2. It would be hard for me to break any of my bad habits. (R)
3. I don't care to know what other people really think of me.
4. I have not always been honest with myself. (R)
5. I always know why I like things.
6. When my emotions are aroused, it biases my thinking. (R)
7. Once I've made up my mind, other people can seldom change my opinion.
8. I am not a safe driver when I exceed the speed limit. (R)
9. I am fully in control of my own fate.
10. It's hard for me to shut off a disturbing thought. (R)
11. I never regret my decisions.
12. I sometimes lose out on things because I can't make up my mind soon enough. (R)
13. The reason I vote is because my vote can make a difference.
14. My parents were not always fair when they punished me. (R)
15. I am a completely rational person.
16. I rarely appreciate criticism. (R)
17. I am very confident of my judgments
18. I have sometimes doubted my ability as a lover. (R)
19. It's all right with me if some people happen to dislike me.
20. I don't always know the reasons why I do the things I do. (R)
21. I sometimes tell lies if I have to. (R)
22. I never cover up my mistakes.
23. There have been occasions when I have taken advantage of someone. (R)
24. I never swear.

25. I sometimes try to get even rather than forgive and forget.
26. I always obey laws, even if I'm unlikely to get caught. (R)
27. I have said something bad about a friend behind his/her back.
28. When I hear people talking privately, I avoid listening. (R)
29. I have received too much change from a salesperson without telling him or her.
30. I always declare everything at customs. (R)
31. When I was young I sometimes stole things.
32. I have never dropped litter on the street. (R)
33. I sometimes drive faster than the speed limit.
34. I never read or look at sexual books, magazines, or websites. (R)
35. I have done things that I don't tell other people about.
36. I never take things that don't belong to me. (R)
37. I have taken sick-leave from work or school even though I wasn't really sick.
38. I have never damaged a library book or store merchandise without reporting it. (R)
39. I have some pretty awful habits.
40. I don't gossip about other people's business. (R)

APPENDIX L
SELF-REPORTED PSYCHOPATHY-III (SRP-III)

SELF-REPORTED PSYCHOPATHY-III (SRP-III)

Williams, K. M., Nathanson, C., & Paulhus, D. L. (2003). Structure and validity of the Self-Report Psychopathy Scale-III in normal populations. *Poster presented at the 111th annual convention of the American Psychological Association*. Toronto, Ontario, Canada.

Items with an (R) were reverse-coded items.

Please rate the degree to which you agree with the following statements about you.

1 = strongly disagree

7 = strongly agree

1. I'm a rebellious person.
2. I'm more tough-minded than other people.
3. I think I could "beat" a lie detector.
4. I have taken illegal drugs (e.g., marijuana, ecstasy).
5. I have never been involved in delinquent gang activity. (R)
6. I have never stolen a truck, car or motorcycle. (R)
7. Most people are wimps.
8. I purposely flatter people to get them on my side.
9. I've often done something dangerous just for the thrill of it.
10. I have tricked someone into giving me money.
11. It tortures me to see an injured animal. (R)
12. I have assaulted a law enforcement official or social worker.
13. I have pretended to be someone else in order to get something.
14. I always plan out my weekly activities. (R)
15. I like to see fist-fights.
16. I'm not tricky or sly. (R)
17. I'd be good at a dangerous job because I make fast decisions.
18. I have never tried to force someone to have sex. (R)
19. My friends would say that I am a warm person. (R)
20. I would get a kick out of 'scamming' someone.
21. I have never attacked someone with the idea of injuring them. (R)
22. I never miss appointments. (R)
23. I avoid horror movies. (R)
24. I trust other people to be honest. (R)
25. I hate high speed driving. (R)
26. I feel so sorry when I see a homeless person. (R)
27. It's fun to see how far you can push people before they get upset.

28. I enjoy doing wild things.
29. I have broken into a building or vehicle in order to steal something or vandalize.
30. I don't bother to keep in touch with my family any more.
31. I find it difficult to manipulate people. (R)
32. I rarely follow the rules.
33. I never cry at movies.
34. I have never been arrested. (R)
35. You should take advantage of other people before they do it to you.
36. I don't enjoy gambling for real money. (R)
37. People sometimes say that I'm cold-hearted.
38. People can usually tell if I am lying. (R)
39. I like to have sex with people I barely know.
40. I love violent sports and movies.
41. Sometimes you have to pretend you like people to get something out of them.
42. I am an impulsive person.
43. I have taken hard drugs (e.g., heroin, cocaine).
44. I'm a soft-hearted person. (R)
45. I can talk people into anything.
46. I never shoplifted from a store. (R)
47. I don't enjoy taking risks. (R)
48. People are too sensitive when I tell them the truth about themselves.
49. I was convicted of a serious crime.
50. Most people tell lies everyday.
51. I keep getting in trouble for the same things over and over.
52. Every now and then I carry a weapon (knife or gun) for protection.
53. People cry way too much at funerals.
54. You can get what you want by telling people what they want to hear.
55. I easily get bored.
56. I never feel guilty over hurting others.
57. I have threatened people into giving me money, clothes, or makeup.
58. A lot of people are "suckers" and can easily be fooled.
59. I admit that I often "mouth off" without thinking.
60. I sometimes dump friends that I don't need any more.
61. I would never step on others to get what I want. (R)
62. I have close friends who served time in prison.
63. I purposely tried to hit someone with the vehicle I was driving.
64. I have violated my parole from prison.

APPENDIX M
DEMOGRAPHICS

DEMOGRAPHICS

What is your current age? _____

With which gender do you self-identify?

Male

Female

With which ethnic group do you self-identify?

European American/White/Caucasian

Hispanic American/Latino(a)

Asian/Asian American/Pacific Islander

African American/Black

Native American/American Indian

Middle Eastern/Arab

Biracial

Other (Please report your ethnicity: _____)

What type of device did you use to complete this study?

Cellular Phone

Tablet (e.g., iPad, Kindle Fire, Samsung Galaxy, etc.)

Laptop/Desktop computer

Other (Please report what device you are currently using: _____)

Did you experience any technological issue with this study (e.g., the webpage took a long time to load, the webpage would not load on your device)?

Yes (if yes, Please explain the issue in the text box provided.)

No

APPENDIX N
DEBRIEFING FORM

DEBRIEFING FORM



College of Social and Behavioral Sciences
Department of Psychology

Debriefing Statement

Thank you for your participation in this study. The true purpose of this study was to examine whether or not data quality could be improved by making people more self-aware of the behavior through the use of honor codes and watchful eyes. In a previous study, we found that a large majority of participants were carelessly responding at CSUSB, which in turn affected our data quality. In this study, we attempted to reduce this careless responding by showing different honor statements. You were either randomly assigned to one of five conditions: You either: (a) only read the honor statement, (b) read-and-retyped the honor statement, (c) read and saw a pair of eyes, (d) saw a pair of eyes, read an honor code, and typed the honor code, or (e) you did not see any of these things.

You will be given your 3 extra credit points no matter how you responded to the survey items.

We request that you do not talk about this study with any of your friends or classmates so that the integrity of the data is not compromised.

Please be assured your name will not be attached in any way to the answers you have provided. Your contributions to our research project are completely anonymous – no one can know that these are your responses. Additionally, no information about your answers will be released to anyone. This is guaranteed and in accordance with ethical and professional codes set by the CSUSB Institutional Review Board and the American Psychological Association.

Furthermore, this has not been an assessment of your ability and/or adequacy. The focus of this research is on all participants as a group (e.g., all college students) and not on individuals. The measures used do not permit meaningful conclusions about individuals. Please contact Dr. Donna M. Garcia at dmgarcia@csusb.edu if you are interested in the results or have any questions.

APPENDIX O
TABLES

Table 1

Correlation Matrix for the Personality Measures

Variables	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
1. Honesty/Humility	<i>.79</i>																	
2. Extraversion	.15**	<i>.86</i>																
3. Agreeableness	.43**	.23**	<i>.81</i>															
4. Conscientiousness	.33**	.33**	.25**	<i>.82</i>														
5. Emotionality	.07	-.17**	-.08	-.06	<i>.79</i>													
6. Openness	.15**	.27**	.17**	.25**	-.14*	<i>.81</i>												
7. Machiavellianism	-.54**	-.29**	-.34**	-.37**	-.13*	-.21**	<i>.75</i>											
8. DOSPERT-Social	.06	.20**	-.06	.07	-.11	.26**	-.07	<i>.63</i>										
9. DOSPERT-Rec.	-.11	.28**	.02	-.04	-.34**	.30**	.04	.32**	<i>.82</i>									
10. DOSPERT-Fin.	-.30**	.06	-.09	-.11	-.17**	.01	.29**	.18**	.35**	<i>.82</i>								
11. DOSPERT-H/S	-.39**	-.08	-.24**	-.28**	-.13*	-.06	.33**	.23**	.32**	.38**	<i>.69</i>							
12. DOSPERT-Ethical	-.53**	-.23**	-.22**	-.41**	-.07	-.10	.50**	.03	.21**	.48**	.53**	<i>.68</i>						
13. RWA	-.05	.09	-.05	.02	.01	-.25**	.08	-.25**	-.06	.06	-.17**	-.09	<i>.76</i>					
14. SDO	-.36**	-.02	-.11	-.19**	-.12*	-.27**	.41**	-.15*	.01	.29**	.20**	.32**	.38**	<i>.93</i>				
15. Self-Esteem	.24**	.66**	.20**	.45**	-.19**	.19**	-.38**	.19**	.12*	-.05	-.16**	-.36**	< .01	-.20**	<i>.92</i>			
16. BIDR-SDE	.24**	.53**	.33**	.46**	-.37**	.25**	-.37**	.23**	.18**	-.04	-.14*	-.30**	-.02	-.18**	.66**	<i>.74</i>		
17. BIDR-IM	.47**	.23**	.41**	.39**	-.14*	.09	-.42**	-.08	-.05	-.15*	-.44**	-.45**	.19**	-.11*	.24**	.40**	<i>.72</i>	
18. S.R. Psychopathy	-.56**	-.03	-.35**	-.37**	-.32**	-.06	.55**	.07	.29**	.44**	.50**	.55**	.07	.50**	-.26**	-.16**	-.41**	<i>.92</i>

Note. N = 305. Personality measures are presented in the order that participants completed them in. S.R. = Self-Reported; Cell = Cellular phone; DOSPERT = Domain-Specific Risk Taking; Rec. =

Recreational; Fin. = Financial; H/S = Health/Safety; RWA = Right Wing Authoritarianism; SDO = Social Dominance Orientation; BIDR-SDE = Balanced Inventory of Desirable Responding Self-

Deception Enhancement; BIDR-IM = Balanced Inventory of Desirable Responding Impression Management. Cronbach alpha's are italicized and on the diagonal. *p < .05, **p < .01.

Table 2

One-Way ANOVAs for the Careless Responding Indices and Participant Engagement based on Honor Code Condition (Hypotheses 1 and 2, Research Questions #1 and #2)

Dependent Variables	<i>F</i>	<i>p</i>	Partial η^2	Honor Code Condition				
				Control (No Honor Code)	Read Only	Type Only	Read and Eyes	Type and Eyes
Instructed Response Items ^a	2.97	.020	.038	1.00 (1.41) _a [0.69, 1.34]	2.25 (3.85) _b [1.42, 3.22]	1.65 (2.50) _{ab} [1.10, 2.38]	1.11 (1.92) _{ab†} [0.71, 1.55]	1.04 (1.71) _{ab†} [0.69, 1.43]
Average LongString ^a	2.61	.036	.034	4.83 (1.18) _{a†} [4.54, 5.16]	9.12 (14.11) _b [6.33, 12.55]	5.68 (3.11) _{ab} [5.01, 6.50]	6.34 (7.79) _{ab} [5.05, 8.08]	6.30 (5.79) _{ab} [5.25, 7.54]
Survey Time ^b	2.09	.083	.030	44.78 (18.87) _a [40.29, 50.23]	39.59 (17.79) _a [34.96, 44.06]	41.46 (17.98) _a [37.06, 46.44]	48.40 (21.95) _{a†} [43.00, 53.85]	40.85 (14.83) _a [36.98, 44.96]
Self-Reported Diligence ^c	1.19	.316	.016	6.00 (0.83) _a [5.80, 6.20]	6.05 (1.12) _a [5.73, 6.34]	6.24 (0.76) _a [6.04, 6.43]	6.21 (0.75) _a [6.01, 6.40]	6.27 (0.76) _a [6.06, 6.47]
Self-Reported Interest ^c	1.99	.097	.026	4.87 (0.94) _a [4.64, 5.10]	5.06 (0.95) _a [4.80, 5.32]	5.11 (1.01) _a [4.85, 5.39]	4.90 (1.10) _a [4.63, 5.16]	5.32 (1.04) _a [5.03, 5.60]

Note. ^a*df* = 4, 300. Means, standard deviations, and 95% bias corrected confidence intervals are based on 1,000 sample bootstrapping analyses. ^b*df* = 4,

274. Survey times reflect scores with 10 percent of the sample trimmed of the right tail of the distribution. Means, standard deviations, and 95% bias corrected confidence intervals are based on 1,000 sample bootstrapping analyses. Survey time is in minutes and timing began after the honor code webpage and ended before the demographics page. ^c*df* = 4, 300. Different subscript letters indicate significant pairwise differences at $p < .05$. A “†” indicates a significant pairwise difference with the read-only condition at $p < .10$.

Table 3

Repeated Measure ANOVAs for the Careless Responding Indices based on Webpage Including Outliers (Research Question #3)

Dependent Variables	<i>F</i>	<i>p</i>	Partial η^2	Webpage		
				Page #1	Page #2	Page #3
Instructed Response Items ^a	0.48	.622	.002	0.45 (0.89) _a	0.49 (0.97) _a	0.47 (0.98) _a
Maximum LongString ^b	10.51	< .001	.033	4.94 (7.24) _a	6.92 (5.60) _b	7.42 (14.12) _b
Page Time ^c	1.96	.156	.006	22.53 (59.14) _a	38.01 (183.46) _a	22.92 (83.25) _a

Note. ^aSphericity was not violated; *df* = 2, 608. ^bGreenhouse-Geisser corrected *F* is reported because sphericity was violated based on Mauchly's test; *df* = 1.51, 457.54. ^cGreenhouse-Geisser corrected *F* is reported because sphericity was violated based on Mauchly's test; *df* = 1.38, 419.46.

Table 4

Repeated Measure ANOVAs for the Careless Responding Indices based on Webpage after Excluding Outliers (Research Question #3)

Dependent Variables	<i>F</i>	<i>p</i>	Partial η^2	Webpage		
				Page #1	Page #2	Page #3
Instructed Response Items ^a	0.58	.559	.002	0.30 (0.59) _a	0.33 (0.63) _a	0.28 (0.60) _a
Maximum LongString ^b	26.75	< .001	.083	4.26 (2.42) _a	6.37 (2.81) _b	5.63 (5.82) _c
Page Time ^c	5.35	.013	.018	18.01 (14.32) _a	21.02 (39.43) _a	14.64 (16.85) _b

Note. ^aSphericity was not violated; *df* = 2, 572. ^bGreenhouse-Geisser corrected *F* is reported because sphericity was violated based on Mauchly's test; *df* = 1.38, 409.49. ^cGreenhouse-Geisser corrected *F* is reported because sphericity was violated based on Mauchly's test; *df* = 1.34, 397.01.

Table 5

One-Way ANOVAs for the Personality Measures based on the Honor Code Conditions (Research Question #4)

Dependent Variables	F	p	Partial η^2	Honor Code Condition				
				Control (No Honor Code)	Read Only	Type Only	Read and Eyes	Type and Eyes
Honesty/Humility ^a	1.66	.158	.022	4.57 (0.82)	4.66 (0.78)	4.52 (0.88)	4.86 (0.73)	4.73 (0.92)
Extraversion ^a	1.68	.154	.022	4.71 (0.84)	4.52 (0.75)	4.45 (1.07)	4.56 (0.94)	4.82 (0.76)
Agreeableness ^a	0.58	.680	.008	4.07 (0.81)	4.18 (0.75)	4.04 (0.92)	4.12 (0.70)	4.24 (0.87)
Conscientiousness ^a	0.50	.734	.007	5.05 (0.81)	4.90 (0.70)	4.92 (0.86)	5.00 (0.73)	5.06 (0.87)
Emotionality ^a	0.28	.891	.004	4.66 (0.88)	4.66 (0.84)	4.77 (0.87)	4.63 (0.78)	4.72 (0.68)
Openness ^a	0.86	.492	.011	4.57 (0.75)	4.47 (0.93)	4.31 (0.84)	4.50 (0.87)	4.37 (0.98)
Machiavellianism ^b	0.71	.583	.009	3.22 (0.67)	3.24 (0.63)	3.36 (0.57)	3.35 (0.60)	3.34 (0.66)
DOSPERT-Social ^a	2.08	.083	.027	4.98 (0.88)	4.54 (1.05)	4.65 (0.90)	4.68 (0.87)	4.77 (0.77)
DOSPERT-Rec. ^a	1.47	.210	.019	4.12 (1.21)	3.59 (1.32)	3.93 (1.31)	3.82 (1.16)	3.78 (1.25)
DOSPERT-Fin. ^a	0.29	.887	.004	2.74 (1.25)	2.61 (1.13)	2.63 (1.01)	2.53 (0.94)	2.63 (1.23)
DOSPERT-H/S ^a	0.57	.686	.008	3.28 (1.11)	3.09 (0.91)	3.28 (0.90)	3.12 (0.92)	3.18 (0.88)
DOSPERT-Ethical ^a	0.93	.445	.012	2.21 (1.01)	2.41 (1.07)	2.47 (1.01)	2.24 (0.81)	2.23 (0.85)
RWA ^b	0.16	.961	.002	3.57 (0.73)	3.57 (0.73)	3.61 (0.77)	3.61 (0.82)	3.67 (0.66)
SDO ^b	1.32	.264	.017	2.47 (1.04)	2.62 (1.03)	2.46 (0.91)	2.21 (0.99)	2.52 (1.15)
Self-Esteem ^b	1.18	.320	.016	5.52 (0.97)	5.48 (1.04)	5.14 (1.34)	5.28 (1.30)	5.41 (0.98)
BIDR-SDE ^a	0.10	.984	.001	4.34 (0.62)	4.32 (0.58)	4.29 (0.75)	4.28 (0.64)	4.34 (0.59)
BIDR-IM ^a	1.57	.182	.021	3.84 (0.71)	3.94 (0.51)	3.90 (0.81)	3.99 (0.62)	4.13 (0.60)
S.R. Psychopathy ^a	0.62	.651	.008	2.79 (0.72)	2.93 (0.71)	2.79 (0.62)	2.76 (0.65)	2.79 (0.65)

Note. ^adf = 4, 300; ^bdf = 4, 299. Personality measures are presented in the order that participants completed them in. DOSPERT = Domain-Specific Risk

Taking; Rec. = Recreational; Fin. = Financial; H/S = Health/Safety; RWA = Right Wing Authoritarianism; SDO = Social Dominance Orientation; BIDR-SDE

= Balanced Inventory of Desirable Responding Self-Deception Enhancement; BIDR-IM = Balanced Inventory of Desirable Responding Impression

Management. S.R. = Self-Reported.

Table 6

Correlation Matrix for the Careless Responding Indices, Participant Engagement, and the Personality Measures (Research Questions #5 and #6)

Variables	Instructed Response Items	Survey Time	Average LongString	S.R. Diligence	S.R. Interest	UseMe
Survey Time	-.01	-				
Average LongString	.44**	-.02	-			
S.R. Diligence	-.38**	-.03	-.35**	.87		
S.R. Interest	-.13*	-.07	-.10	.40**	.80	
UseMe	-.30**	.04	-.23**	.35**	.18**	-
Technological Issues	.01	-.03	-.03	-.01	.04	-.14*
Comp. [0] vs. Cell [1]	.26**	-.03	> .01	-.20**	-.11	-.16**
Comp. [0] vs. Tablet [1]	-.04	-.06	.01	-.07	-.01	-.10
Honesty/Humility	-.18**	-.01	-.10	.20**	.25**	.04
Extraversion	-.07	-.03	-.02	.13*	.21**	.02
Agreeableness	-.09	-.04	-.01	.08	.17**	.05
Conscientiousness	-.19**	-.04	-.13*	.26**	.26**	.05
Emotionality	-.20**	-.02	-.08	.01	.10	.05
Openness	-.01	-.01	-.04	.10	.18**	-.01
Machiavellianism	.23**	.03	.15*	-.27**	-.32**	-.08
DOSPERT-Social	-.09	.06	-.17**	.18**	.10	.06
DOSPERT-Rec.	.10	> -.01	.01	> -.01	.03	-.06
DOSPERT-Fin.	.22**	> -.01	.11*	-.21**	-.02	-.13
DOSPERT-H/S	.10	.04	< .01	-.11*	-.16**	-.06
DOSPERT-Ethical	.29**	.05	.14*	-.31**	-.16**	-.24**
RWA	.05	.07	.11	-.13*	-.03	-.04
SDO	.26**	-.01	.21**	-.38**	-.19**	-.10
Self-Esteem	-.18**	-.06	-.14*	.28**	.21**	.08
BIDR-SDE	-.13*	-.07	-.09	.25**	.20**	.05
BIDR-IM	-.01	-.08	.02	.07	.21**	-.02
S.R. Psychopathy	.33**	< .01	.31**	-.39**	-.25**	-.15**

Note. $N = 305$. Personality measures are presented in the order that participants completed them in. UseMe and Technological Issues: No

= 0, Yes = 1; Comp. = Laptop/Desktop computer; Cell = Cellular phone; DOSPERT = Domain-Specific Risk Taking; Rec. = Recreational;

Fin. = Financial; H/S = Health/Safety; RWA = Right Wing Authoritarianism; SDO = Social Dominance Orientation; BIDR-SDE = Balanced

Inventory of Desirable Responding Self-Deception Enhancement; BIDR-IM = Balanced Inventory of Desirable Responding Impression

Management. S.R. = Self-Reported. Cronbach alpha's are italicized and on the diagonal. * $p < .05$, ** $p < .01$.

Table 7

Correlation Matrix for Instructed Response Items and Personality Measures for the Entire Sample, Control Condition, and Read Only Condition and Fisher r-to-z Transformation Values between the Control and Read Only Conditions

Variables	Instructed Response Items				Observed Power for Fisher r-to-z
	Entire Sample ^a	Control (No Honor Code) ^b	Read Only Condition ^c	Fisher r-to-z transformations	
Honesty/Humility	-.18**	-.23	-.25	0.11	.06
Extraversion	-.07	-.26*	-.12	-0.78	.19
Agreeableness	-.09	-.19	-.10	-0.49	.47
Conscientiousness	-.19**	-.21	-.32*	0.64	.16
Emotionality	-.20**	.02	-.38**	2.25*	.73
Openness	-.01	-.13	-.02	-0.59	.15
Machiavellianism	.23**	.16	.39**	-1.34	.38
DOSPERT-Social	-.09	-.03	.11	-0.75	.19
DOSPERT-Rec.	.10	.01	.14	-0.70	.17
DOSPERT-Fin.	.22**	.22	.37**	-0.88	.22
DOSPERT-H/S	.10	.11	.03	0.43	.11
DOSPERT-Ethical	.29**	.17	.37**	-1.16	.31
RWA	.05	.18	-.03	1.13	.30
SDO	.26**	.33**	.23	0.58	.14
Self-Esteem	-.18**	-.22	-.37**	0.88	.22
BIDR-SDE	-.13*	-.24	-.16	-0.45	.12
BIDR-IM	-.01	-.06	.09	-0.81	.20
S.R. Psychopathy	.33**	.18	.44**	-1.56	.46

Note. ^a*N* = 305. ^b*n* = 62. ^c*n* = 58-59. Personality measures are presented in the order that participants completed them in. DOSPERT = Domain-Specific

Risk Taking; Rec. = Recreational; Fin. = Financial; H/S = Health/Safety; RWA = Right Wing Authoritarianism; SDO = Social Dominance Orientation;

BIDR-SDE = Balanced Inventory of Desirable Responding Self-Deception Enhancement; BIDR-IM = Balanced Inventory of Desirable Responding

Impression Management. S.R. = Self-Reported. Cronbach alpha's are italicized and on the diagonal. **p* < .05, ***p* < .01.

APPENDIX P
FIGURES

CALIFORNIA STATE UNIVERSITY
SAN BERNARDINO

DEPARTMENT OF
PSYCHOLOGY

For this study you will complete eight different personality measures broken up onto three web pages.



Completing this study honestly is part of CSUSB's academic ethics code. Therefore, I agree to read all instructions and respond to all of the survey items honestly. I also will not provide responses to survey items that do not reflect my true qualities. Providing good quality answers is important to the field of psychology and I wish to do my part by answering truthfully.

>>

Figure 1. Screenshot of the type-and-eyes honor code page.

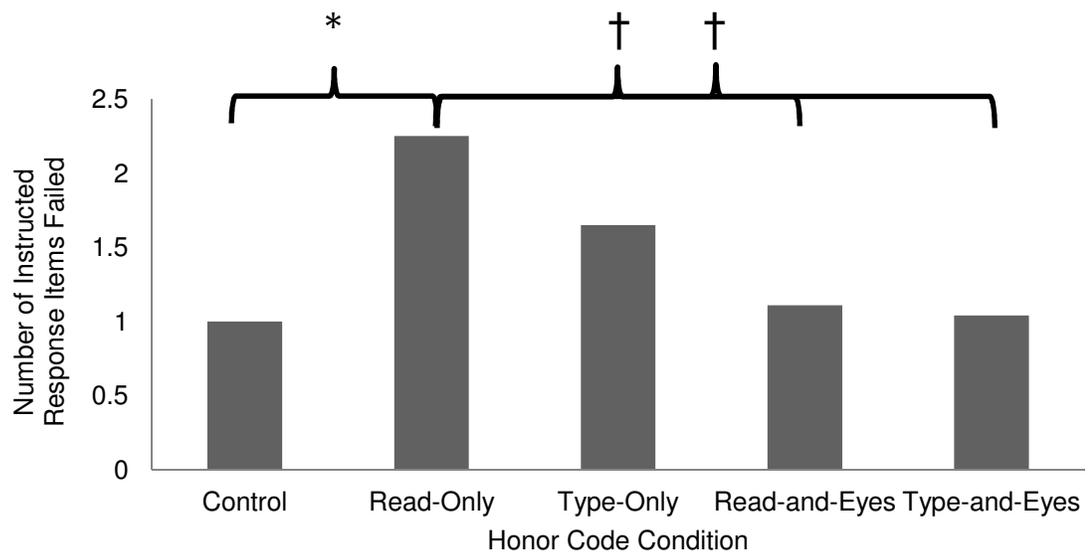


Figure 2. Number of instructed response items failed by honor code condition. * $p < .05$; † $p < .10$.

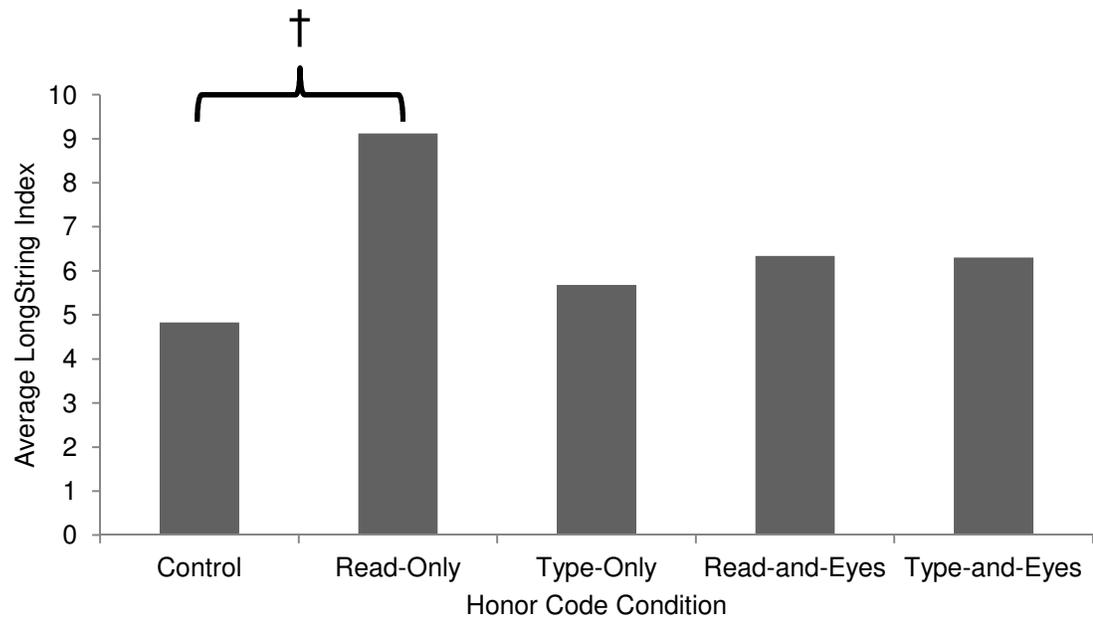


Figure 3. Average LongString index score by honor code condition. † $p < .10$.

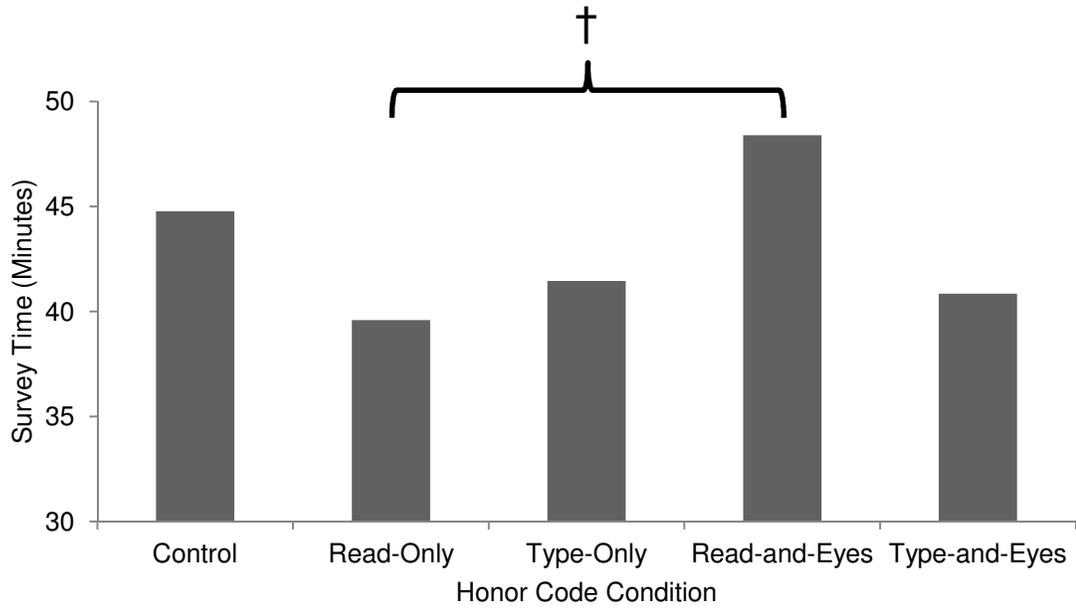


Figure 4. Survey Time in minutes by honor code condition. † $p < .10$.

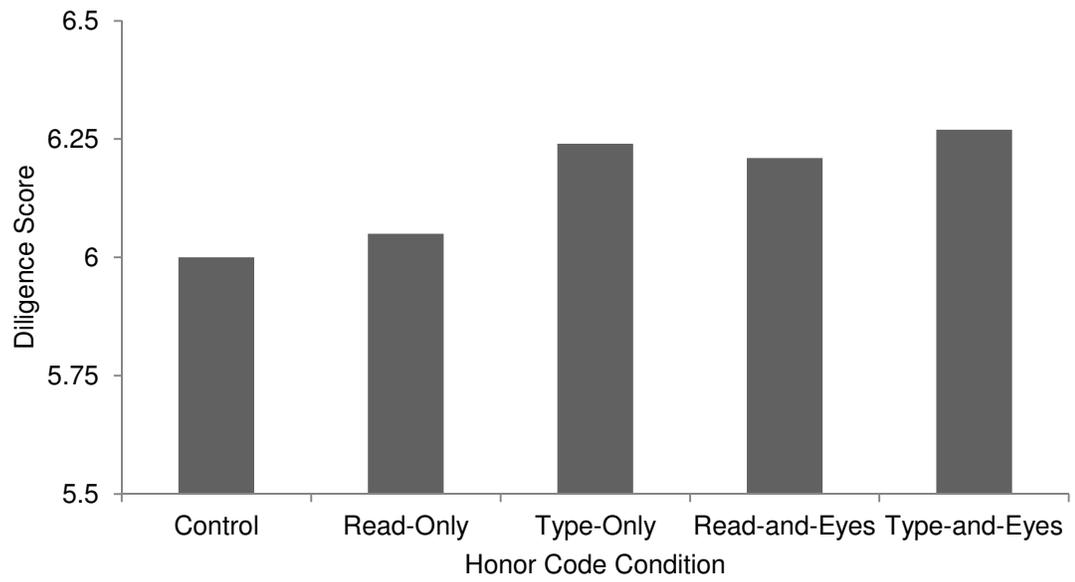


Figure 5. Diligence score by honor code condition.

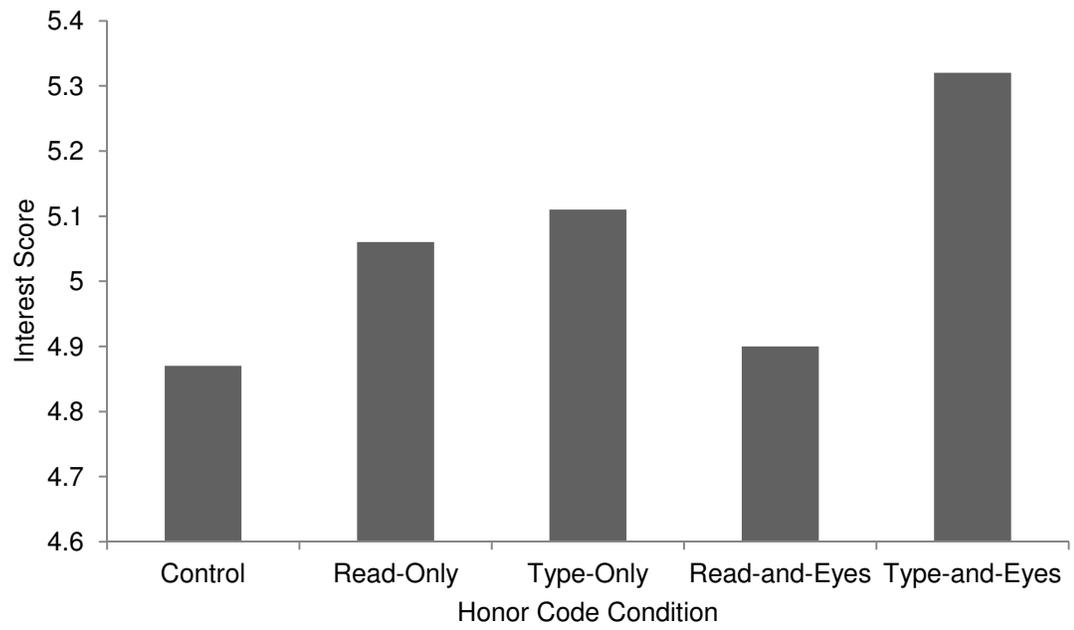


Figure 6. Interest score by honor code condition.

REFERENCES

- Altemeyer, B. (1998). The other "Authoritarian Personality." *Advances in Experimental Social Psychology*, *30*, 47–92. doi:10.1016/S0065-2601(08)60382-2
- Ariely, D. (2012). *The honest truth about dishonesty: How we lie to everyone—especially ourselves*. London, UK: HarperCollins Publishers.
- Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440. doi:10.1037//0022-3514.83.6.1423
- Baer, R. A., Ballenger, J., Berry, D. T. R., & Wetter, M. W. (1997). Detection of random responding on the MMPI-A. *Journal of Personality Assessment*, *68*, 139–151. doi:10.1207/s15327752jpa6801_11
- Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (2000). Detecting and discriminating between random responding and overreporting on the MMPI–A. *Journal of Personality Assessment*, *72*, 308–320. doi:10.1207/S15327752JP720213
- Barber, L. K., Barnes, C. M., & Carlson, K. D. (2013). Random and systematic error effects of insomnia on survey behavior. *Organizational Research Methods*, *16*, 616–649. doi:10.1177/1094428113493120
- Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*, 412–414. doi:10.1098/rsbl.2006.0509

- Beach, D. A. (1989). Identifying the random responder. *Journal of Psychology*, *123*, 101–103. doi:10.1080/00223980.1989.10542966
- Beaman, A. L., Klentz, B., Diener, E., & Svanum, S. (1979). Self-awareness and transgression in children: Two field studies. *Journal of Personality and Social Psychology*, *37*, 1835–1846. doi:10.1037/0022-3514.37.10.1835
- Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*, 800–813. doi:10.3758/s13428-011-0081-0
- Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*, *4*, 340–345. doi:10.1037/1040-3590.4.3.340
- Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior*, *23*, 463–477. doi:10.1016/j.chb.2004.10.020
- Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. New York, NY: Academic Press.
- Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, *90*, 125–144. doi:10.1348/000712699161189
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on*

Psychological Science, 6, 3–5. doi:10.1177/1745691610393980

Caruso, E. M., Vohs, K. D., Baxter, B., & Waytz, A. (2013). Mere exposure to money increases endorsement of free-market systems and social inequality. *Journal of Experimental Psychology: General*, 142, 301–316. doi:10.1037/a0029288

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29, 2156–2160. doi:10.1016/j.chb.2013.05.009

Cassileth, B. R., Zupkis, R. V., Sutton-Smith, K., & March, V. (1980). Informed consent - why are its goals imperfectly realized? *The New England Journal of Medicine*, 302, 896–900. doi:10.1056/NEJM198004173021605

Chapman, L. J., & Chapman, J. P. (1983). *Infrequency Scale*. Unpublished test.

Chiaburu, D. S., Huang, J. L., Hutchins, H. M., & Gardner, R. G. (2013).

Trainees' perceived knowledge gain unrelated to the training domain: The joint action of impression management and motives. *International Journal of Training and Development*, 18, 37–52. doi:10.1111/ijtd.12021

Cho, E., & Allen, T. D. (2012). Relationship between work interference with family and parent–child interactive behavior: Can guilt help? *Journal of Vocational Behavior*, 80, 276–287. doi:10.1016/j.jvb.2011.12.002

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York, NY: Academic Press.

- Costa, P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). *The SAGE handbook of personality theory and assessment*, 2179-2198.
- Curran, P. G., Kotrba, L., & Denison, D. (2010, April). Careless responding in surveys: Can traditional techniques be applied to real world settings? *Paper presented at the 25th annual conference of the Society for Industrial/ Organizational Psychology*. Atlanta, GA.
- Dijksterhuis, A., & Van Knippenberg, A. (2000). Behavioral indecision: Effects of self-focus on automatic behavior. *Social Cognition*, 18, 55–74.
doi:10.1521/soco.2000.18.1.55
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16, 1–23. doi:10.1207/S15327043HUP1601_1
- Ehlers, C., Greene-Shortridge, T. M., Weekley, J. A., & Zajack, M. D. (2009, April). *The exploration of statistical methods in detecting random responding*. Paper presented at the 24th annual meeting of the Society for Industrial/Organizational Psychology, New Orleans, LA.
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, 63, 591–601. doi:10.1037/0003-066X.63.7.591
- Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human*

- Behavior*, 32, 172–178. doi:10.1016/j.evolhumbehav.2010.10.006
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
doi:10.3758/BRM.41.4.1149
- Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, 72, 892–913.
doi:10.1093/poq/nfn059
- Gervais, W. M., & Norenzayan, A. (2012). Like a camera in the sky? Thinking about God increases public self-awareness and socially desirable responding. *Journal of Experimental Social Psychology*, 48, 298–302.
doi:10.1016/j.jesp.2011.09.006
- Gino, F., & Pierce, L. (2009). The abundance effect: Unethical behavior in the presence of wealth. *Organizational Behavior and Human Decision Processes*, 109, 142–155. doi:10.1016/j.obhdp.2009.03.003
- Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes*, 115, 191–203.
doi:10.1016/j.obhdp.2011.03.001
- Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms

that affect the consistency of responses to synonym and antonym pairs.

Journal of Personality and Social Psychology, 48, 82–98.

doi:10.1037/0022-3514.48.1.82

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59, 93–104.

doi:10.1037/0003-066X.59.2.93

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, 68, 155–165. doi:10.1037/0022-006X.68.1.155

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26, 245–256. doi:10.1016/j.evolhumbehav.2005.01.002

Heatherton, T. F., & Wyland, C. L. (2003). Assessing self-esteem. In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures* (pp. 219–233). Washington, D.C.: American Psychological Association.

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, 21, 360–373. doi:10.1177/0894439303253985

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–83.

doi:10.1017/S0140525X0999152X

Howell, D. C. (2010). *Statistical methods for psychology (7th edition)*. Belmont, CA: Cengage Wadsworth.

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*, 99–114. doi:10.1007/s10869-011-9231-8

Jenni, K. E., & Loewenstein, G. (1997). Explaining the “Identifiable Victim Effect.” *Journal of Risk and Uncertainty, 14*, 235–257.

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

Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, and Computers, 31*, 433–438. doi:10.3758/BF03200723

Kurtz, J. E., & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment, 76*, 315–332.

doi:10.1207/S15327752JPA7602_12

Lee, K., & Ashton, M. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39*, 329–358.

doi:10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and Narcissism

in the Five-Factor Model and the HEXACO model of personality structure. *Personality and Individual Differences*, *38*, 1571–1582.

doi:10.1016/j.paid.2004.09.016

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. doi:10.1016/j.jrp.2013.09.008

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*, 1–23.

doi:10.3758/s13428-011-0124-6

Mažar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*, 633–644. doi:10.1509/jmkr.45.6.633

Mažar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy and Marketing*, *25*, 117–126.

doi:10.1509/jppm.25.1.117

McCabe, D. L., & Trevino, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *Journal of Higher Education*, *64*, 522–538.

doi:10.2307/2959991

McCabe, D. L., & Trevino, L. K. (1997). Individual and contextual influences on academic dishonesty: A multicampus investigation. *Research in Higher Education*, *38*, 379–396.

McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2002). Honor codes and other

contextual influences on academic integrity: A replication and extension to modified honor code settings. *Research in Higher Education*, *43*, 357–378.

McKay, A. S., Mussel, P., & Kaufman, J. C. (2014). *Bogus items and instructed response items: An investigation of their effectiveness for identifying careless responders*. Manuscript in preparation.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437–455. doi:10.1037/a0028085

Miller, E. T., Neal, D. J., Roberts, L. J., Baer, J. S., Cressler, S. O., Metrik, J., & Marlatt, G. A. (2002). Test-retest reliability of alcohol measures: Is there a difference between Internet-based assessment and traditional methods? *Psychology of Addictive Behaviors*, *16*, 56–63. doi:10.1037//0893-164X.16.1.56

Nichols, D. S., Greene, R. L., & Schmolck, P. (1986). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, *45*, 239–250. doi:10.1002/1097-4679(198903)45:2%3C239::AID-JCLP2270450210%3E3.0.CO;2-1

O'Dell, J. W. (1971). Method for detecting random answers on personality questionnaires. *Journal of Applied Psychology*, *55*, 380–383. doi:10.1037/h0031473

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional

manipulation checks: Detecting satisficing to increase statistical power.

Journal of Experimental Social Psychology, 45, 867–872.

doi:10.1016/j.jesp.2009.03.009

Osborne, J. W., & Blanchard, M. R. (2010). Random responding from participants is a threat to the validity of social science research results.

Frontiers in Psychology, 1. doi:10.3389/fpsyg.2010.00220

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P.

Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of*

personality and social psychological attitudes (pp. 17–59). San Diego, CA:

Academic Press.

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality:

Narcissism, Machiavellianism, and psychopathy. *Journal of Research in*

Personality, 36, 556–563. doi:10.1016/S0092-6566(02)00505-6

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social Dominance Orientation: A personality variable predicting social and political attitudes.

Journal of Personality and Social Psychology, 67, 741–763.

doi:10.1037/0022-3514.67.4.741

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological*

Bulletin, 114, 510–532. doi:10.1037/0033-2909.114.3.510

Rentfrow, P. J. (2010). Statewide differences in personality: Toward a

psychological geography of the United States. *American Psychologist*, 65,

548–558. doi:10.1037/a0018194

- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754–775. doi:10.1037//0021-9010.84.5.754
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*, 160–177. doi:10.1177/1073191103252349
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO-PI-R: Development and initial validation. *Journal of Personality Assessment, 68*, 127–138. doi:10.1207/s15327752jpa6801_10
- Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89*, 623–642. doi:10.1037/0022-3514.89.4.623
- Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9*, 367–373. doi:10.1177/014662168500900405
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93–105. doi:10.1037/0003-066X.54.2.93
- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to

- study clinical populations. *Clinical Psychological Science*, 1, 213–220.
doi:10.1177/2167702612469015
- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: Self-preservation through moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37, 330–349. doi:10.1177/0146167211398138
- Shu, L. L., Mažar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, 109, 15197–15200. doi:10.1073/pnas.1209746109
- Stanton, J. M. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology*, 51, 709–725. doi:10.1111/j.1744-6570.1998.tb00259.x
- Stanton, J. M., & Rogelberg, S. G. (2001). Using Internet/Intranet web pages to collect organizational research data. *Organizational Research Methods*, 4, 200–217. doi:10.1177/109442810143002
- Van Zandt, T. (2002). Analysis of response time distributions. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology volume 4* (pp. 461–516). New York, NY: John Wiley & Sons.
- Vincent, L. C., Emich, K. J., & Goncalo, J. A. (2013). Stretching the moral gray zone: Positive affect, moral disengagement, and dishonesty. *Psychological Science*, 24, 595–599. doi:10.1177/0956797612458806

- Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*, 263–290. doi:10.1002/bdm.414
- Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods, 18*, 53–70. doi:10.1037/a0031607
- Williams, K. M., Nathanson, C., & Paulhus, D. L. (2003). Structure and validity of the Self-Report Psychopathy Scale-III in normal populations. In *Poster presented at the 111th annual convention of the American Psychological Association*. Toronto, Ontario, Canada.
- Wiltermuth, S. S., & Gino, F. (2013). “I’ll have one of each”: How separating rewards into (meaningless) categories increases motivation. *Journal of Personality and Social Psychology, 104*, 1–13. doi:10.1037/a0030835
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory Factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28*, 186–191. doi:10.1007/s10862-005-9004-7
- Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology, 22*, 51–68. doi:10.1002/acp.1331
- Zakrisson, I. (2005). Construction of a short version of the Right-Wing Authoritarianism (RWA) scale. *Personality and Individual Differences, 39*,

863–872. doi:10.1016/j.paid.2005.02.026

Zhong, C.-B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police:

Darkness increases dishonesty and self-interested behavior.

Psychological Science, 21, 311–314. doi:10.1177/0956797609360754

Zook, A. (1985). On measurement of Machiavellianism. *Psychological Reports*,

57, 11–12. doi:10.2466/pr0.1985.57.3.982