10-5-2017

Assessing the Effects of Survey Instructions and Physical Attractiveness on Careless Responding in Online Surveys

Carolyn M. Rauti
University of Windsor

Assessing the Effects of Survey Instructions and Physical Attractiveness on Careless Responding in Online Surveys

by
Carolyn M. Rauti

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the Requirements for
the Degree of Master of Arts at the
University of Windsor

Windsor, Ontario, Canada

Assessing the Effects of Survey Instructions and Physical Attractiveness on Careless Responding in Online Surveys

by
Carolyn M. Rauti

APPROVED BY:

_____
R. Arnold
Department of Sociology, Anthropology and Criminology

_____
P. Fritz
Department of Psychology

_____
D. Jackson, Advisor
Department of Psychology

September 14, 2017

Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

The current study explored the effects of survey instructions (basic, warning, feedback) and survey administrator appearance (invisible administrator, higher attractiveness, lower attractiveness) on careless responding in online surveys. Undergraduate students (N = 527) were randomly assigned to one of nine experimental conditions and completed an online survey regarding personality, attitudes and experiences in University. Three two-way ANOVAs and one two-way ANCOVA were used in this study. Conscientiousness was used as a covariate and careless responding behavior was measured by total survey response time, response consistency, response patterns, and self-reported carelessness. The findings indicated that higher levels of conscientiousness were related to lower levels of self-reported carelessness, and that survey instructions and survey administrator appearance do have some influence on careless responding behavior.

Table of Contents

List of Tables

List of Figures

List of Appendices

Assessing the Effects of Survey Instructions and Physical Attractiveness on Careless Responding in Online Surveys

CHAPTER I: Introduction

Advances in technology have increased the use of online surveys as a means to collect data in research. Online survey administration offers several advantages as it is cost effective and time efficient, provides easier access to larger samples, and is convenient for both researchers and respondents (Riggle, Rostosky, & Reedy, 2005; Shwarz, 1999; Ward, Clark, Zabriskle, & Morris, 2014; Wright, 2005). Despite these advantages, this mode of survey administration is not without its drawbacks. Previous research suggests that data obtained from online surveys are susceptible to the subtle yet harmful effects of suboptimal responses from respondents who are inattentive or distracted. Suboptimal responses may also come from respondents who are unmotivated to comply with survey instructions, interpret item content correctly, or provide thoughtful and accurate responses (Berinsky, Margolis, & Sances, 2013; Huang, Curran, Keeney, Poposki, DeShon, 2012). In recent years, researchers have acted to better understand and measure suboptimal responses that result from careless responding behavior.

Careless responding has been defined as intentionally or unintentionally responding to survey items in a way that does not accurately reflect ones' true feelings or beliefs (Ward & Pond, 2015). In the literature, it has often been referred to as inattentive responding (McGrath et al., 2010), insufficient-effort responding (Bowling, Huang, Bragg, Khazon, & Blackmore, 2016) and satisficing (Barge & Gehlbach, 2012). Estimates of the prevalence of careless responding appear to vary by study, ranging from 3-46% of data (e.g., Curran et al., 2010; Johnson, 2005; Meade & Craig, 2012) and may be more pervasive than many researchers realize. Careless responding poses a threat to

data quality and inferences drawn from research, and therefore it is crucial to create viable solutions to minimize it.

Careless Responding Detection Methods

To avoid its harmful effects, past research has extensively focused on careless responding detection methods (e.g., Akbulut, 2015; Huang et al., 2012; Huang, Bowling, Liu, & Lu, 2014; Meade & Craig, 2012), and from this, several asserted effective screening indices have been proposed. There is no single detection method to identify all possible types of careless responses; however, researchers typically screen for carelessness by inserting specialized items into the survey and by evaluating respondents' survey performance after data collection.

Specialized items inserted into the survey may include self-report items in which respondents are asked to indicate their level of attentiveness during survey completion, whether the responses provided reflect true feelings and/or beliefs, and whether the responses provided are of adequate quality for researcher use (Ward & Pond, 2015). It has been suggested that self-report items as such are generally effective in detecting careless responses as respondents tend to answer these items honestly; however, this type of indicator is insufficient on its own (Meade & Craig, 2012). Similar to this approach, researchers can insert specialized "trap questions" often referred to as instructional manipulation check (IMCs) items in their surveys. A typical IMC is a survey item that instructs participants to provide an unconventional response in place of an intuitively correct answer (Hauser, Sunderrajan, Natarajan, & Schwarz, 2016). IMCs require respondents to pay close attention to answer the item correctly, thus incorrect responses are used as indications that respondents failed to pay close attention and were careless.

Miller and Baker-Prewitt (2009) note that failure on trap questions is highly correlated with satisficing; however, using such items demonstrates a lack of respect for survey respondents as these items seem trivial to those who are fully paying attention. It has also been argued that use of trap questions may degrade data quality (Vanette, 2017) as it may induce a Hawthorne effect or social desirability bias (i.e., change in responses due to feeling of being watched), and therefore should be avoided.

The second general method of careless responding detection includes procedures that measure respondents' survey performance after data collection. Indices such as response time, response consistency, and response patterns are commonly used in data cleaning procedures (e.g., DeSimone, Harms, & DeSimone, 2015; Meade & Craig, 2012; Ward & Pond, 2015). The response time approach assumes that careless responders will have shortened response times on individual survey items and in total duration relative to non-careless responders. Huang et al. (2012) note that although variations in reading speed and item length make cutoff scores difficult to justify, it should take participants at least 2 seconds per item to respond. Shorter response times may indicate that respondents skimmed or rushed through the survey without fully cognitively processing the content before selecting a response option. Built-in software timing features can be used to indicate whether participants rushed or skipped items by assessing the amount of time spent on each individual item, on an individual page of items, or on the total survey (Barge & Gelbach, 2012; DeSimone, Harms, & DeSimone, 2015; Robinson-Cimpian, 2014).

Response consistency can be assessed by examining whether respondents provided similar responses to survey items of similar content. Inconsistent responses to

similar meaning items are thought to indicate carelessness (Lucas & Baird, 2005; Meade & Craig, 2012). A commonly used response consistency indicator is the Even-Odd Consistency measure (e.g., Johnson, 2005; Meade & Craig, 2012), which divides the even items from the odd items using a unidimensional scale. Within-person correlations across the pairs of items are then computed and compared. Small within-person correlations across the subsets of paired items are thought to indicate careless responding (Ward & Pond, 2015).

The response patterns approach allows researchers to identify the extent to which respondents selected a single response option. If survey items are randomly ordered and some items are reverse scored, it would not be possible to consistently choose a single response option and doing so would likely indicate that participants provided inaccurate responses. To assess response patterns, the longest string of consecutive items in which respondents have selected the same response option is computed and a maximum long string value is assigned to each respondent (Huang et al., 2012; Johnson, 2005; Meade & Craig, 2012). Maximum long string values on a measure with *k* items ranges from 1 to *k*-1, and larger values are used as an indication of greater carelessness.

It is important to note that although these detection methods can screen data for careless responses, data cleaning procedures can never be completely accurate and it has been suggested that removal of respondents' data is problematic as it reduces sample size in a nonrandom way, artificially shapes the sample distribution, limits the generalizability of findings and narrows the implications of the study (Maniaci & Rogge, 2014; Ward & Pond, 2015). To improve data quality, it is not only necessary to identify effective methods to minimize careless responding, it is also crucial to understand why individuals

engage in this pattern of responding in the first place.

Explanations for Careless Responding

Past research suggests that several factors are at play when understanding why individuals carelessly respond. For instance, levels of motivation and attention needed for careful responding may reflect an individual's personality traits and behavioral characteristics. Individuals high in conscientiousness, a personality trait characterized as being thorough, careful and vigilant (Richardson & Abraham, 2009) are likely to be more careful when responding to survey items based on defining characteristics of their personality. Because responding carefully to a questionnaire requires attention to detail and willingness to follow instructions, conscientious participants may naturally respond carefully due to their general tendency to be attentive and compliant (Meade & Pappalardo, 2013). A recent study conducted by Bowling et al. (2016) supported this notion as conscientiousness was negatively related to indices measuring insufficient-effort responding.

In contrast to conscientiousness, individuals high in impulsivity, a trait characterized by a tendency to display behavior of little to no forethought or reflection, tend to be more careless when completing tasks. Past research has noted that impulsiveness is positively related to inattention (Colledge & Blair, 2001), lack of focus on a task (Bechara, Damasio, & Damasio, 2000), and greater focus on short-term gains such as obtaining immediate reward (Diekhof et al., 2012). These findings may suggest that participants who score higher in impulsivity may be less attentive when responding to a questionnaire or desire to complete the questionnaire quickly.

In addition to respondents' personality traits related to carelessness, concern over respondents' motivation and attentiveness is likely intensified as survey research has moved to an online format. Past research suggests that administrators of online surveys have forfeited the supervision and control that they had when overseeing traditional paper-pencil surveys (Huang et al., 2014; Meade & Craig, 2011). The absence of direct interaction or social exchange between the researcher and respondent (Gehlbach & Barge, 2012; Johnson, 2005) as well as the increased likelihood of multitasking and environmental distractors (Zwarun & Hall, 2014) may increase respondent inattentiveness.

Researchers have also investigated fatigue effects associated with cognitive processing (i.e., taking mental short cuts and putting less effort into a task) that may be related to survey responding. The cognitive demands required for completing a survey such as reading items thoroughly and responding accurately (Weijters, De Beuckelaer, & Baumgartner, 2014) is thought to relate to careless responding if individuals fail to cognitively process the items that they are responding to (Berinsky, Margolis, Sances, 2013). Theories of satisficing (e.g., Krosnick, 1991; Simon, 1957) have also been used to understand respondents' cognitive processing and exerted cognitive effort that may produce suboptimal responses (Barge & Gelbach, 2012; Tourangeau, Rips, & Rasinsky, 2000). The satisficing phenomenon refers to taking mental shortcuts rather than considering a full range of options when responding to survey items. Respondents may satisfice by selecting the first option rather than the best option (Hauser et al., 2016), and in extreme cases may select responses at random (Krosnick, 1991). Johnson (2005) noted that satisficing may occur in unsupervised online surveys due to the social distance

between the researcher and respondent, and perceived anonymity and ease of survey submission online.

In relation to fatigue effects, the length of the survey is thought to relate to carelessness as respondents may experience fatigue or boredom when lengthy questionnaires (e.g., inventories that contain several hundreds of items) exceed ones' attention span. Because careful responding to lengthy surveys require high levels of sustained attention, lengthy surveys may result in respondents' desire to skip or rush through survey items without fully processing the content (Maniaci & Rogge, 2014). Levels of engagement as well as motivation to spend time thinking about questions before responding, especially in lengthy questionnaires, are thought to decrease with surveys on topics that are trivial or nonrelevant to respondents (Holbrook, Krosnick, Moore, & Tourangeau, 2007).

These explanations may suggest that the prevalence of careless responding in online surveys is associated with survey design characteristics. While controlling for personality variables that are related to carelessness, improving online survey methodology by including design features that increase respondents' level of engagement and attentiveness may prove to be crucial for reducing careless responding behavior. Several studies have attempted to examine the effects of survey instructions on responding behavior; and, to a lesser extent, past researchers have investigated the effects of online survey administrator presence to mimic the social connection between the researcher and respondent as a means to influence online survey responding behavior.

Survey Instructions

Past research suggests that the type of survey instructions that respondents are presented with prior to completing an online survey can influence responding behavior. A large body of literature has focused on *warning* instructions that hint at punitive consequences for carelessness and a smaller proportion of research has focused on *feedback* instructions that give participants feedback on some aspect of performance. As discussed below, several studies have compared the effectiveness of these types of instructions to *basic/normal* (control) instructions.

Warning messages seek to reduce the likelihood of satisficing by increasing participants' motivation to provide an accurate answer to survey items (Clifford & Jerrit, 2015; Krosnick, 2000). These findings are explained by operant conditioning theories (Skinner, 1938) which suggest that punishment is effective in behavior modification. That is, warning respondents of potential consequences for low-quality responses may increase attentiveness presumably to avoid the occurrence of such consequences. A study conducted by Huang et al. (2012) tested this by comparing the effects of normal instructions (simply asking participants for honesty and informing them that there are no right or wrong answers) to warning instructions (telling participants advanced statistical control procedures will detect insufficient responding and result in loss of participation credit) on several careless responding indices. The results from this study showed that those who were given the warning instructions provided fewer careless responses compared to those who were given normal instructions. Further, respondents in the warning condition had greater consistency and reliability in their responses to survey items.

Clifford and Jerrit (2015) tested the effects of four different types of warning

messages compared to a control group and found that three of the four warning messages indicated greater attentiveness than the control group, and one of the four warning messages indicated greater engagement than the control group. Meade and Craig (2012) also found that warning survey instructions decreased the prevalence of careless responding and participants in the warning condition self-reported a greater level of attentiveness while completing the survey. These findings were later replicated by Ward and Pond (2015) who found that respondents given warning instructions had significantly smaller maximum long string values than those who were given normal instructions.

Past research has noted that offering an incentive such as evaluative feedback on a task can improve ones' attentiveness and performance. As indicated by Kluger and DeNisi (1996), feedback intervention (FI) theory proposes that when offered feedback on task performance, respondents are more attentive to their actions and this shift in attentiveness tends to improve their task performance. Northcraft, Schmidt and Ashford (2011) tested the FI theoretical model and found that individuals invested more time and effort and tended to perform better on tasks for which performance feedback was available. Gosling, Vazire, Srivastava, and John (2004) noted that providing feedback appeals to individuals' desire for self-insight, and participants are motivated to answer honestly to receive accurate feedback about themselves and/or their performance. Ward and Pond (2015) examined the effects of promising performance feedback on careless responding in their online survey where they compared the survey responses of participants given normal instructions to responses from participants given feedback instructions (telling participants they will receive feedback on the quality of their responses). The authors found that on average, participants in the feedback condition took

longer to answer items and self-reported greater data quality, suggesting that participants were more attentive and careful when responding to the survey items.

Studies examining the effects of warning and feedback survey instructions on careless responding have only compared their effectiveness to basic (control group) instructions. Thus, whereas both warning instructions and instructions providing evaluative feedback have shown to be effective in shaping responding behavior, it is currently unknown if one of the two is more effective in reducing careless responding in a student sample. Exploring whether one type of message is more effective may partially provide a more effective option for obtaining better data quality.

Survey Administrator Presence

Previous literature has suggested that careless responding in online surveys may, in part, be due to the absence of social interaction between the researcher and respondent (Johnson, 2005). Behrend and Foster-Thompson (2011) noted that inducing a perceived social interaction between the survey administrator and respondent may increase respondents' accountability and attentiveness during survey completion due to an induced perception of supervision. Ward and Pond (2015) examined this notion and tested whether the presence of a virtual survey administrator influenced participants' responding behavior. In this study, the virtual survey administrator conditions consisted of an animated slightly moving circular shape which appeared from the beginning of the survey until completion, or a virtual human survey administrator with movements such as blinking and breathing. These conditions were compared to a control group with no visible survey administrator. The authors found that respondents in the virtual human condition scored lower on a multivariate composite of careless responding compared to

those in the control group and animated shape conditions. Further, there was a significant interaction between virtual presence and instructional messages. Posthoc analyses indicated that those exposed to the virtual human and the warning message provided significantly fewer careless responses. Although these findings suggest that incorporating a virtual researcher into the design of an online survey may increase participant attentiveness; a more advanced method for including a survey administrator may indicate improved results. It was of interest to assess whether including a more realistic connection between the survey administrator and respondent and whether the physical characteristics of the survey administrator have a greater influence on respondents' attentiveness during the completion of an online survey.

Physical Appearance

Characteristics, such as one's physical appearance, serve as an important evaluative cue in person perception and influences how one is treated by others (Agnew, 1984; Dion & Berschield, 1974; Sigall & Ostrove, 1975). Although many claim that "beauty is in the eye of the beholder," some evidence suggests (e.g., Coetzee, Greeff, Stevens, & Perrett, 2014) that there are within- and cross-cultural agreement in facial attractiveness preferences (i.e., shiny hair, youthful or flawless skin, and symmetrical facial features). Anecdotally speaking, the mass amount of commercials advertising skincare products for clear and youthful skin, or haircare products for healthy, shiny hair, as well as the surge in cosmetic procedures used to enhance one's appearance, show some evidence to support this claim.

Research pitting individuals who vary in attractiveness against one another have consistently shown that physically attractive individuals are evaluated more positively on

a wide range of personal characteristics (e.g., friendliness, intelligence and warmth) whereas unattractive individuals are evaluated more negatively on these same characteristics (Dion, Berschield, & Walster, 1975; Lorenzo, Biesanz, & Human, 2010; Lucker, Beane, & Helmreich, 1981). The stereotypical belief which often assumes "what is beautiful is good" is commonly referred to as a *halo effect*.

Consistent with attractiveness stereotypes in other domains, studies have shown that students rate attractive teachers as more competent, more motivating, and better at stimulating learning (Chaikin, Gillen, Derlega, Heinen & Wilson, 1978). A professor's level of attractiveness has also shown to influence students' level of engagement and learning outcomes (Gurung & Vespia, 2007; Riniolo, Johnson, Sherman, & Misso, 2006). That is, compared to unattractive professors, students who have attractive professors are likely to exhibit higher levels of engagement in class and are more likely to earn better grades as a result. An experimental study conducted by Westfall (2015) demonstrated that, with all else being equal, students assigned to a condition with an attractive teacher performed better on a recall test than students assigned to a condition with an unattractive teacher.

Past literature has suggested that physical appearance influences observers' visual attention span. Researchers that have examined this attractiveness-visual attention phenomenon have indicated that individuals look at faces higher in attractiveness for a longer period of time than faces lower in attractiveness (Aharon et al., 2001; Langlois, Ritter, Roggman & Vaughn, 1991) and pay more attention to those deemed attractive (Sui & Liu, 2009). Westfall (2015) suggested that more attention may be paid to attractive individuals because physically attractive people tend to be perceived more

positively and perceivers may consider physically attractive individuals more worthy of attention. Literature on persuasion tends to support the notion that physically attractive individuals have some degree of control over observers' behaviors as people are more likely to pay attention to an attractive speaker, and this increases the odds that a message given by an attractive speaker will be remembered (Perloff, 2014). Thus, as previous literature suggests that physical appearance influences engagement and attention, it was of interest to test whether these findings extend to survey administrator appearance exerting influence on respondents' survey responding behaviors.

The Current Study

The intent of the current research was to better understand whether certain combinations of survey design features (types of survey instructions and survey administrator appearance) can reduce careless responding in online surveys. To control for traits thought to be associated with careless behavior, this research examined whether personality characteristics of conscientiousness and impulsivity were related to careless responding measures. Careless responding was measured by four separate indices including total survey response time, response consistency, response patterns, and a self-reported measure of carelessness.

Research Questions

Three research questions were of interest in each of the analyses conducted.

Question 1: Overall, is one type of instructional message more effective in reducing careless responding as measured by careless responding indices?

Question 2: Does the survey administrator's appearance influence participants' responding behaviors as measured by careless responding indices?

Question 3: Is there an interaction between survey instructions and survey administrator appearance on the careless responding indices?

Outcome Expectations

Hypothesis 1: Although studies indicate that both incentives and warnings of punishment are effective in short-term behavior modification (Balliet, Mulder, & Van Lange, 2011; Kubanek, Snyder, & Abrams, 2015), there is not a clear consensus on which strategy is more effective. However, given that the sample used in this research was undergraduate students who participated in the study to obtain a course bonus credit, it is likely that the warning instructions would be more effective in influencing responding behavior compared to the performance feedback instructions. Presumably, undergraduate students would be more likely to follow instructions to avoid possible penalization, especially when it is associated with their final grade in a course.

Hypothesis 2: Based on previous research suggesting that individuals higher in physical attractiveness influence observers' behaviors (e.g., Gurung & Vespia, 2007; Riniolo, Johnson, Sherman & Misso, 2006), it was expected that the survey administrator higher in attractiveness would influence participants' responding by increasing attentiveness and engagement. Specifically, it was expected that participants in the higher attractiveness conditions would show lower levels of carelessness compared to participants in the other two conditions.

Hypothesis 3: Based on evidence indicating a significant interaction between message type and inclusion of a virtual researcher (i.e., Ward & Pond, 2015), an interaction between the independent variables in the current study was expected. Because a significant interaction between the threatening message-type and inclusion of virtual

human researcher was found in multivariate measure of careless responding, it was anticipated that participants in the warning and higher attractive condition would provide fewer careless responses in comparison to participants in all other conditions.

CHAPTER II: Methodology

Participants

The total sample consisted of 527 undergraduate students from the University of Windsor. Cell sizes per experimental condition ranged from 54 to 63 participants due to random assignment. The majority of the sample were female (81.2%), and the average age of participants was 22 years old (Range = 17- 58, Median = 20). More participants were currently in their fourth year or higher (28.8%), followed by third (27.9%), second (23.9%) and first (19.5%) year of study. Table 1 presents the demographic statistics.

Participants were recruited through the psychology department's participant pool system which is an online recruitment tool where participants registered in the pool must be enrolled in at least one undergraduate psychology or business course. Studies that are listed in the participant pool are presented in a random order and participants can select the studies in which they wish to participate. Participants were not informed of the true intent of this research and instead were told that the purpose of the study was to examine personality characteristics and student attitudes and behavior in University. Those who participated were sent a web-link to one of nine versions of the online survey where they provided consent to participate, completed questionnaires, were debriefed, and entered their email address to receive one bonus point that could be allocated to a participating course they were enrolled in. Data collection took place in the winter and intersession semesters of 2017.

Table 1

Participant Demographics

Variable		*n*	*%*
Age			
	M 21.65		
	SD 4.93		
Gender			
	Male	99	18.8
	Female	428	81.2
First year of study			
	Yes	114	21.6
	No	399	75.7
	Missing Response	14	2.7
Taken courses prior to attending the University			
	Yes	9	1.7
	No	457	86.7
	Missing Response	61	11.6
Program of study			
	FAHSS	288	54.6
	Business	41	7.8
	Human Kinetics	36	6.8
	Math and Sciences	67	12.7
	Education	7	1.3
	Nursing	30	5.7
	Engineering	7	1.3
	Inter-Faculty	31	5.9
	Other	20	3.9
Ethnicity			
	Caucasian/White	328	62.1
	African American/Canadian	41	7.8
	Asian	32	6.1
	Middle Eastern	60	11.4
	Hispanic/Latin	7	1.3
	Native Canadian	3	0.6
	Inter-Racial	20	3.8
	Other	36	6.8
Student status			
	Canadian	503	95.3
	American	2	0.4
	International	20	3.8
	Missing Response	2	0.4
Year of study			
	1	102	19.3
	2	125	23.7

3	146	27.7
4 or more	151	28.6
Missing Response	3	0.6

Note. FAHSS = Faculty of Arts, Humanities, and Social Sciences

Study Design

A 3x3 between-subjects experimental design was used to assess the effects of survey instructions (basic, warning, feedback) and survey administrator appearance (invisible administrator, higher attractiveness, lower attractiveness) on careless responding. Participants were randomly assigned to one of nine experimental conditions (described below) where respondents were exposed to some combination of instructional message and survey administrator appearance. All participants completed the same sequence of surveys. Careless responding was measured by four indices including total response time, response consistency, response patterns, and a self-reported measure of carelessness.

Experimental Conditions

Instructional message type

Participants were given one of three types of survey instructions (these instructions were adapted from Ward & Pond, 2015). To ensure the instructions were understood, participants were required to type out the instructions they received in an open text box before they could move to the next page and respond to survey items.

Basic instructions. Participants in this condition served as the control group for the instructions manipulation. The basic instructions stated *"Welcome to our study. During this study, you will be asked to complete several questionnaires based on personality, attitudes, and behaviors in University. Your honest and thoughtful responses are important to us and to this study."*

Warning instructions. The warning instructions began with the basic instructions but included a subsequent message stating *"...To ensure the quality of survey data, your responses will be subject to sophisticated statistical control methods. Responding carelessly will be flagged as low-quality data and may result in reduced bonus credit."*

Feedback instructions. The feedback instructions began with the basic instructions but included a subsequent message stating *"...You will receive feedback based on the quality of your responses and whether we can use the information you have provided to us, upon completion of the survey."*

Administrator Appearance

The survey administrator's appearance was displayed to participants in one of three ways. In the two conditions where the administrator was visible, participants could see the administrator's face and upper body. In the condition where the survey administrator was not visible, a black box appeared.

Invisible administrator. Participants in this condition served as the control group for the appearance manipulation. In this condition, participants could not see the administrator but could hear the administrator providing survey instructions.

Higher attractiveness. The appearance of the survey administrator was manipulated using makeup. Participants in the higher attractiveness conditions viewed a video of the survey administrator providing survey instructions.

Lower attractiveness. The appearance of the survey administrator was manipulated through the misuse of makeup. Participants in the lower attractiveness conditions viewed a video of the survey administrator providing survey instructions.

Procedure

Survey administrator interviews. Prior to the study, recruitment for a female actress was advertised to students in the Dramatic Arts program at the University of Windsor. The researcher of this study and a small group of graduate students held brief interviews with each of the five candidates. During the interviewing process, candidates were informed about the nature of the research study, their expected role, and compensation. Upon agreement amongst those present in the interview, one candidate was employed to act as the survey administrator. The selected candidate was considered high in attractiveness yet could be made to appear less attractive with the misuse of makeup. Further, the selected candidate was a fourth-year undergraduate student and had more acting experience in comparison to the other four candidates.

Instructional videos. The instructional videos were filmed on the University of Windsor campus in the fall semester of 2016. To assist in creating the videos, both a make-up artist and videographer were employed. The manipulation of the survey administrator's appearance for both the higher and lower attractive conditions were approved by the small group of those present during the filming session.

Online survey. Nine versions of the online survey were created through FluidSurveys.com. The survey began with a consent form followed by a video of survey instructions with an open text box asking participants to type out their understanding of the instructions they received. This was mandatory to move forward in the survey and responses were analyzed to ensure that participants understood the instructions given; those who answered this item incorrectly were discarded from analyses. Following the survey instructions page, there were seven questionnaires, debriefing information, and a

separate page for participants to enter their email address to receive compensation.

Survey Content

Several measures were used in this study, some for the purposes of controlling for personality characteristics related to carelessness, and some for measuring the degree of careless responding within experimental conditions. The measures that were used are described below.

The Big-Five Inventory (BFI). The BFI (Goldberg, 1993) is a 44-item inventory that measures the Big Five personality factors: extroversion, agreeableness, conscientiousness, neuroticism and openness. Items on this measure include: "I see myself as someone who is talkative," "I see myself as someone who can be somewhat careless," and "I see myself as someone who worries a lot." Participants respond to the items using a 5-point Likert scale, ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). In past research, the BFI has demonstrated good reliability with an average Cronbach's alpha coefficient of 0.85 (Soto & John, 2009). In the current study, conscientiousness was the only subscale of interest.

Baratt's Impulsiveness Scale (BIS-11). The BIS-11 (Patton, Stanford & Barratt, 1995) is a 30-item inventory used to measure the personality and behavioral constructs of impulsiveness and nonimpulsiveness (for reverse scored items). The inventory measures three dimensions of impulsiveness labelled as attentional (task-focus, intrusive thoughts and racing thoughts), motor (acting on spur of the moment) and nonplanning (careful thinking and planning). Items on this measure include: "I plan tasks carefully," "I am a careful thinker," and "I don't pay attention." Participants respond to items on a 4-point Likert scale from 1(*rarely/never*) to 4(*almost always/always*). In past research, the BIS-

11 has demonstrated good internal consistency, with an average Cronbach's alpha coefficient of 0.80 (Reise, Moore, Sabb, Brown, & London, 2014).

Academic Stress Scale. The Academic Stress Scale (Kohn & Frazier, 1986) is a 35-item measure of stress experienced by students. Items on this scale include common academic events such as buying books, having excessive homework, and speaking in class. Participants rate each event on a scale from 0-100. An event considered to be as stressful as taking an examination is to be rated as 50. If the event is less stressful than taking an examination it is to be rated between 0-49, and if the event is considered more stressful than taking an examination it is to be rated between 51-100. Past research (e.g., Burnett & Fanshawe, 1996; Kohn & Frazer, 1986) has indicated excellent internal reliability, with an average Chronbach's alpha coefficient of 0.92.

Academic Well-Being. The Academic Well-Being scale (Chambel & Curral, 2005) is a 10-item scale that is used to measure student burnout and engagement based on academic work demands and control. Items on this scale include both positive and negative emotions and behaviors including feeling depressed, feeling tense, and feeling anxious. Participants respond to items on a 7-point Likert scale from 1 (*never*) to 7 (*all the time*) where higher scores are thought to indicate higher levels of well-being. The scale has demonstrated good reliability in the past, with a Chronbach's alpha value of 0.90 (Chambel & Curral 2005).

Psychological Entitlement Questionnaire. The Psychological Entitlement Scale (Campbell, Bonacci, Shelton, Exline, & Bushman, 2004) is a 9-item measure of general psychological entitlement. Items include: "Great things should come to me," "If I were on the Titanic, I would deserve to be on the first life boat!" and "Things should go my

way." Participants respond to items using a 7-point Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). This scale has shown to be reliable with a Chronbach's alpha coefficient of 0.87 (Campbell et al., 2004).

Academic Entitlement Questionnaire. The Academic Entitlement Questionnaire (Jackson, Singleton-Jackson, Frey, & Mclellan, 2013) is a 61-item multi-dimensional measure of academic entitlement. This scale measures seven domains including general entitlement, reward for effort, accommodation, responsibility avoidance, customer orientation, customer service expectations, and grade haggling. Items on this scale include: "I should never fail an assignment I put effort into," "Great academic success should just come to me," and "A professor should modify course requirements to help me achieve a better grade." Participants respond to items using a 7-point Likert scale ranging from 1 (*strongly disagree)* to 7 (*strongly agree*). Cronbach's alpha from previous versions of this questionnaire suggest good to excellent internal consistency with coefficients ranging from 0.75 to 0.95 (Reinhardt, 2011).

Demographics. A 19-item demographic questionnaire was used to gather data on participants' age, gender, ethnicity, year of study, program major, GPA, studying habits (e.g., number of hours per week studying alone) and parenting variables (e.g., country of origin and household income).

Manipulation Check Items. The survey included three manipulation check items. All participants were asked: "To what extent did the survey instructions that you received influence your level of attentiveness when responding to the survey items." This item was rated on a scale from 1 (*not at all*) to 7 (*very much*). Participants in the higher and lower attractiveness conditions responded to two items regarding their perception of

the survey administrator's physical appearance. The items included: "Please rate the survey administrator's physical appearance on a scale from 1 (*not at all physically attractive*) to 10 (*very physically attractive*), and "Would you generally consider the survey administrator to be lower in physical attractiveness, average, or higher in physical attractiveness?" It was expected that responses to these items would be related (i.e., a participant who rated the survey administrator's appearance as 7 out of 10, should have rated the survey administrator as higher in attractiveness when responding to the subsequent item).

Self-report carelessness indicator. Participants were asked to respond to a single item measuring self-reported carelessness: "To what extent do you think your responses reflects your true sentiments and are of sufficient quality for researchers to use?" This item was rated 1 (*very poor quality*) to 7 (*very good quality*).

CHAPTER III: Results

Data from nine experimental conditions were combined and coded into one large dataset. All analyses were conducted using SAS version 9.3 and SPSS version 24.

Analysis of Manipulation Check Items

An independent samples *t*-test was conducted to compare the attractiveness ratings between participants assigned to the higher and lower attractiveness conditions. The results indicated that when asked to rate the appearance of the survey administrator from 1 (*not at all physically attractive*) to 10 (*very physically attractive*), participants assigned to the higher attractive conditions rated the administrator higher in attractiveness ($n = 176$, $M = 7.64$, $SD = 1.35$) compared to those assigned to the lower attractiveness conditions ($n = 162$, $M = 6.61$, $SD = 1.89$). This difference was statically significant,

$t(348) = 5.90$, $p < .001$, Cohen's d = .60. Similarly, a Chi-square (χ^2) independence test indicated significant differences between the two conditions when asked to categorize the survey administrator's appearance by unattractive, average, or attractive, $\chi^2(2, n = 352) = 29.98$, $p < .001$. An odd's ratio calculation indicated that participants in the higher attractiveness conditions were 2.21 times more likely to rate the survey administrator's appearance as attractive compared to those in the lower attractiveness conditions.

When participants were asked to indicate the extent to which the survey instructions they received influenced their level of attentiveness to survey items, those who received the warning instructions reported the highest influence ($n = 174$, $M = 4.91$, $SD = 1.66$), followed feedback instructions ($n = 163$, $M = 4.07$, $SD = 1.77$), and basic instructions ($n = 169$, $M = 3.97$, $SD = 1.90$). A one-way ANOVA revealed statistically significant differences between the three groups, $F(2, 521) = 14.40$, $p < .001$, $\omega^2 = .04$. Bonferonni posthoc tests indicated those given warning instructions rated this item significantly higher than those given the basic ($p < .001$) and feedback instructions ($p < .001$); however, ratings between the basic and feedback groups did not significantly differ from each other ($p = 1.00$).

Main Analyses

Analysis #1: Response Time

Strategy. The total time taken to complete the survey was recorded from Fluildsurveys.com software and response times were recoded into minutes and seconds in SPSS. Shorter response times were thought to indicate careless responding. It was expected that conscientiousness and impulsivity would be related to response time; however, correlation analysis indicated that neither conscientiousness nor subscales

measuring impulsivity were significantly correlated with total response time. A two-way ANOVA was conducted to assess whether survey instructions and administrator appearance influenced participants' response time. Simple main effect analyses were used to interpret the significant findings.

Assumptions. An analysis of *z* score calculations indicated that 15 cases exceeded a cut-off value above |2.5|, a value used as the general rule of thumb for determining outliers (Fields, 2013). These response times were substantially higher than the other scores (with values ranging from 279 mins and 52 secs to 1,407 mins and 24 secs) and likely were from individuals who left their survey browser open for an extended period of time. These cases were discarded from subsequent analyses to avoid altering the mean response time in experimental conditions. After outliers were removed, this analysis included data from 512 respondents and cell sizes per experimental condition ranged from 51 to 63 cases.

Univariate normality was assessed both statistically and using graphical methods. Skewness and kurtosis values of each experimental condition indicated non-normal distributions. Shapiro-Wilk's test of normality also indicated violations of this assumption with p values < .05 in each condition. Histograms illustrated a positively skewed distribution in each condition, and normal q-q plots illustrated deviations of observed data from a normal distribution. A log transformation was computed on the response time variable due to non-normality.

Levene's test of equality of error variances indicated homogeneity of variance within experimental conditions, F (8, 503) = 1.93, p < .06. Further, it was assumed that

observations were independent as respondents completed this survey from their own computer in varied locations.

Findings. A two-way ANOVA was conducted to examine differences in response time between each experimental condition. The means and standard deviations of the experimental conditions are located in Table 2. The results indicated a significant interaction between survey administrator appearance and survey instructions on response time, $F(4, 503) = 2.98$, $p < .05$, $\omega^2 = .005$. The results from the ANOVA are found in Table 3.

Table 2

Mean (Standard Deviation) of Response Time per Experimental Condition

	Invisible administrator	Higher attractiveness	Lower attractiveness	Total
Basic Instructions				
M (*SD*)	3.28 (.41)	3.36 (.67)	3.03 (.17)	3.25 (.51)
n	56	56	63	175
Warning Instructions				
M (*SD*)	3.20 (.41)	3.35 (.40)	3.37 (.49)	3.31 (.44)
n	52	60	51	163
Feedback Instructions				
M (*SD*)	3.25 (.36)	3.31 (.52)	3.44 (.53)	3.33 (.48)
n	59	63	52	174
Total				
M (*SD*)	3.35 (.39)	3.34 (.54)	3.30 (.49)	
N	167	179	166	

Table 3

ANOVA Results with Response Time as the Dependent Variable

Source	SS	df	MS	*F*	*p*	ω^2
Instructions	.50	2	.25	1.10	.33	
Administrator Appearance	.83	2	.42	1.85	.16	
Instructions x Appearance	2.68	4	.67	2.90	.02	.005

Simple main effects analysis revealed significant differences in response time between the invisible administrator and higher attractiveness conditions and between the higher attractiveness and lower attractiveness conditions (p values < .05) when given basic instructions. The results also indicated significant differences in response time between the invisible administrator and lower attractiveness conditions (p < .05) when given feedback instructions. These findings are illustrated in Figure 1. Although a significant interaction was hypothesized, these results did not support the hypothesis that participants given warning instructions with a higher attractive survey administrator would have longer response times (i.e., would be more careful when responding to survey items) compared to the other conditions.

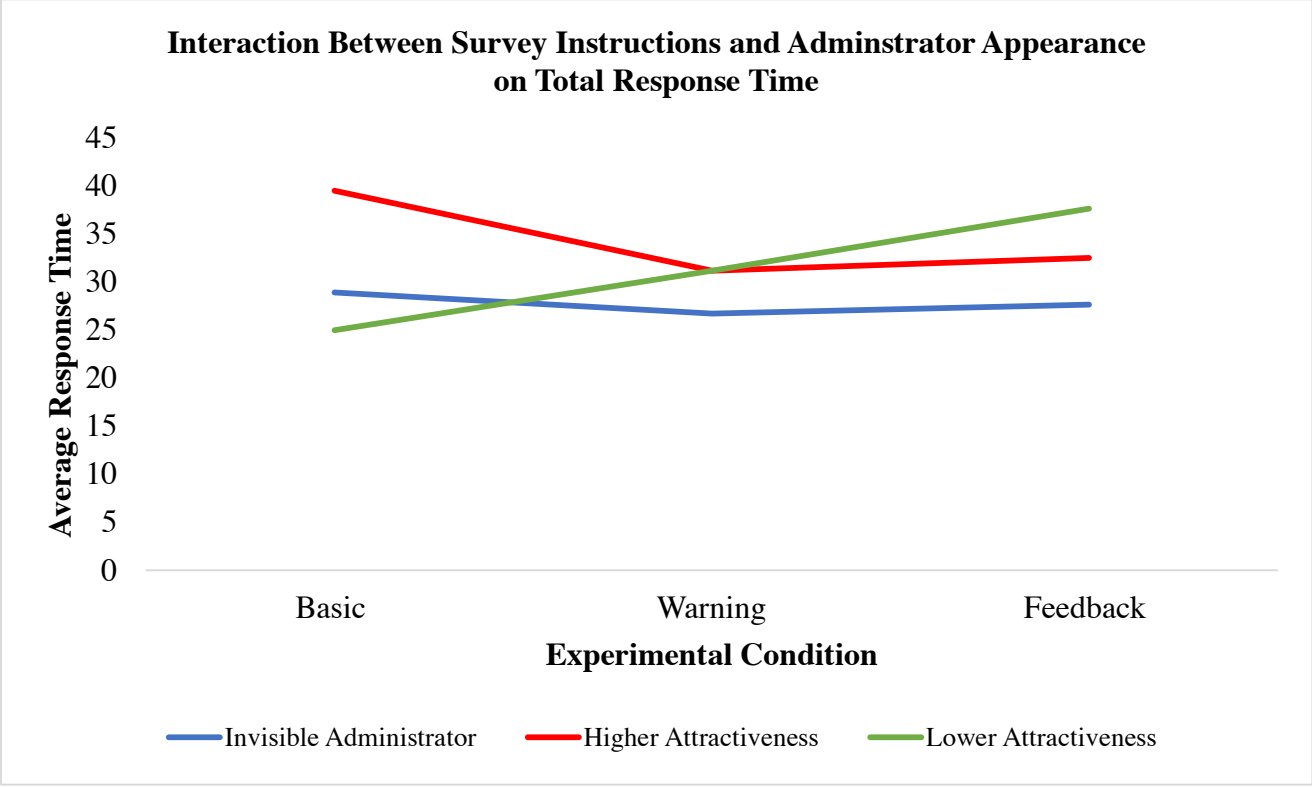

Figure 1. Significant differences in appearance levels were found when given basic instructions and feedback instructions.

Analysis #2: Response Consistency

Strategy. The Academic Stress scale was used to compute the Even-Odd consistency indicator[1]. This scale was split into two subscales of the even and odd numbered items. A within-person correlation was computed for the even and odd pairs of items where values can range from -1 to 1; lower values were thought to indicate careless responding. The within-person correlation value was used as the outcome variable.

Although it was expected that conscientiousness and impulsivity would be related to participants' response consistency, correlation analysis showed that neither conscientiousness nor scales measuring impulsivity were significantly correlated with this variable (p values > .05). A two-way ANOVA was conducted to assess the effects of survey instructions and administrator appearance on response consistency.

Assumptions. An analysis of z score calculations indicated that 7 cases exceeded a cut off value of |2.5|. These extreme scores ranged in value from -.22 to -.49. Given that the intent of this study was to assess respondents' degree of carelessness, these cases were not treated as extreme scores and were retained in the analysis. It should be noted that removal of these cases did not change the findings. Data from 527 participants were used in this analysis with experimental conditions ranging from 54 to 63 cases.

Statistical and graphical methods indicated that the assumption of univariate normality was met in most experimental conditions. Skewness and kurtosis values of each experimental condition did not exceed +/- 2 and +/- 3, respectively, and visual interpretation of histograms and q-q plots illustrated relatively normal distributions.

[1] Item 35 from the Academic Stress scale was left out of the even-odd consistency calculations

Shapiro Wilk's test of normality also indicated univariate normality with the exception of conditions of basic instructions with no survey administrator visible (p = .04), and basic instructions with the lower attractiveness (p = .03).

Levene's test of equality of error variance indicated homogeneity of variance amongst experimental conditions $F(8, 518) = .41$, $p = .91$. Further, it was assumed that observations were independent as respondents completed this survey from their own computer in varied locations.

Findings. A two-way ANOVA was conducted to examine differences in response consistency between each experimental condition. The response consistency values ranged from -.49 to .89; the means and standard deviations of experimental conditions are shown in Table 4. Contrary to hypotheses, results indicated that survey instructions and survey administrator appearance did not significantly affect response consistency, nor was there an interaction between these two variables (p values > .05). The findings from the ANOVA are displayed in Table 5.

Table 4

Mean (Standard Deviation) of Response Consistency per Experimental Condition

	Invisible administrator	Higher attractiveness	Lower attractiveness	Total
Basic instructions				
M (*SD*)	.39 (.22)	.39 (.24)	.32 (.22)	.37 (.23)
n	59	58	63	180
Warning instructions				
M (*SD*)	.36 (.21)	.34 (.23)	.40 (.22)	.36 (.22)
n	55	60	55	170
Feedback instructions				
M (*SD*)	.36 (.20)	.40 (.21)	.38 (.22)	.38 (.21)
n	60	63	54	177
Total				
M (*SD*)	.37 (.21)	.38 (.22)	.37 (.22)	
N	174	181	172	

Table 5

ANOVA Results with Response Consistency as the Dependent Variable

Source	SS	df	MS	*F*	*p*
Instructions	.02	2	.01	.24	.79
Administrator appearance	.01	2	.003	.06	.94
Instructions x Appearance	.30	4	.08	1.56	.18

Analysis #3: Response Patterns

Strategy. The scales included in the maximim long string calculation were the Academic Well-Being Scale, Psychological Entitlement Questionnaire, and Academic Entitlement Questionnaire. These three scales summed to a total of 80 items. Maximum long string values indicated the maximum number of consecutively repeated responses. Maximum long string values could range from 0-79 and larger values were thought to indicate careless responding. A maximum long string value was computed for each participant. Correlation analysis indicated conscientiousness and scales measuring impulsivity were not significantly related to response patterns. A two-way ANOVA was conducted to assess whether survey instructions and administrator appearance influenced response patterns.

Assumptions. An analysis of *z* score calculations indicated that 11 cases exceeded a cut-off value above |2.5|. These values were substantially higher than the average long string value ($M = 5.69$, $SD = 8.40$) with values ranging from 27 to 79. Interestingly, the 11 cases with extreme long string values were those given basic instructions ($n = 7$) and feedback instructions ($n = 4$). As mentioned previously, given that the intent of this study was to assess degree of carelessness, these cases were not treated as outliers and were retained in the analysis. It should be noted that removal of

these cases did not change the findings. This analysis included data from 527 respondents and cell sizes per experimental condition ranged from 54 to 63 cases.

Univariate normality was assessed both statistically and using graphical methods. Skewness and kurtosis values of each condition indicated several non-normal distributions as values exceeded +/- 2 and +/- 3, respectively. Shapiro-Wilk's test of normality also indicated violations of this assumption with $p < .05$ in each condition. Histograms indicated positively skewed distributions, and normal q-q plots illustrated deviations of observed data from a normal distribution in each condition. Normality violations were likely due to the fact extreme scores were retained; however, ANOVA is robust to non-normal data and the positively skewed distributions consistent in each condition, as well as the large sample size should help alleviate problems associated with this violated assumption.

Levene's test of equality of error variances failed to indicate homogeneity of variance within experimental conditions, F (8, 518) = 2.34, $p < .05$, and analysis of group variances showed that the largest group variance was more than 4 times greater than the smallest group variance. It should be noted that ANOVA is generally robust to homogeneity of variance violations when sample sizes are approximately equal. Further, it was assumed that observations were independent as respondents completed this survey from their own computer in varied locations.

Findings. Descriptive analysis showed that maximum long string values ranged from 1 to 79. The means and standard deviations of the experimental conditions are located in Table 6. Contrary to hypotheses, the results from the two-way ANOVA indicated that survey instructions and administrator appearance did not significantly

affect respondents' response patterns, nor was there an interaction between these two variables (p values > .05). The results from the ANOVA are located in Table 7.

Table 6

Mean (Standard Deviation) of Response Patterns per Experimental Condition

	Invisible administrator	Higher attractiveness	Lower attractiveness	Total
Basic instructions				
M (*SD*)	6.08 (9.55)	5.91 (10.52)	7.79 (13.16)	6.63 (11.21)
n	59	58	63	180
Warning instructions				
M (*SD*)	5.05 (3.25)	4.85 (3.46)	4.22 (2.28)	4.71 (3.06)
n	55	60	55	170
Feedback instructions				
M (*SD*)	6.17 (11.11)	4.87 (3.80)	6.06 (9.19)	5.67 (8.5)
n	60	63	54	177
Total				
M (*SD*)	5.79 (8.73)	5.20 (6.65)	6.10 (9.64)	
N	174	181	172	

Table 7

ANOVA Results with Response Patterns as the Dependent Variable

Source	SS	df	MS	*F*	*p*
Instructions	31.82	2	155.91	2.21	.11
Administrator appearance	60.78	2	30.39	.43	.65
Instructions x Appearance	37.55	4	37.55	.53	.71

Analysis #4: Self-Reported Carelessness

Strategy. The single item assessed participants' self-reported carelessness. This item was reverse worded; lower scores on this item indicated a greater degree of self-reported carelessness. Correlation analysis indicated that conscientiousness was significantly related to self-report carelessness ($r = .19$, $p < .001$); however, scales measuring impulsivity were not. An ANCOVA was conducted to examine whether

survey instructions and survey administrator appearance influenced participants' perception of their data quality, while controlling for conscientiousness.

Assumptions. An analysis of *z* score calculations indicated 13 cases that had exceeded a cut off value of |2.5|. These extreme cases ranged from 1-3 and although were considerably lower than the average response on this item ($M = 6.07$, $SD = 1.02$), these cases were retained for analyses. Data from 527 participants were used in this analysis with experimental conditions ranging from 51 to 61 cases.

Tests of univariate normality indicated non-normality. Although the skewness and kurtosis values of each experimental condition did not exceed +/- 2 and +/- 3, respectively, histograms illustrated negatively skewed distributions for each condition and q-q plots showed deviations from normal distributions. Further, Shapiro Wilk's test of normality indicated univariate normality was violated in each condition (p values < .001).

Levene's test of equality of error variances indicated homogeneity of variance within experimental conditions, $F(8, 493) = .52$, $p = .84$, and it was assumed that observations were independent as respondents completed this survey from their own computer in varied locations. Linearity between the covariate and outcome variable at levels of the independent variables were assessed through visual inspection of a matrix scatterplot. The matrix scatterplot illustrated elliptical shapes in each experimental condition indicating a linear relationship between the covariate and outcome variable. Analysis of homogeneity of regression slopes indicated this assumption was met as p values associated with each combination of interactions between the independent variables and covariate were above a value of .05.

Findings. Descriptive analysis showed that scores on the self-reported carelessness item ranged from 1 (*very poor quality*) to 7 (*very high quality*); the means and standard deviations of experimental conditions are located in Table 8. As shown in Table 9, the results from a two-way ANCOVA indicated that, while controlling for conscientiousness, there was a significant main effect for survey instructions on self-reported carelessness, $F(2, 492) = 5.93$, $p = .003$, $\omega^2 = .0004$.

Bonferonni posthoc analysis indicated that scores on the self-reported carelessness indicator significantly differed between the conditions of warning and feedback instructions ($p = .001$). However, basic and warning instructions, and basic and feedback instructions did not significantly differ from each other ($p > .05$). This finding is illustrated in Figure 2. The results from this analysis support the hypothesis that warning messages would be more effective in reducing carelessness when compared to basic and feedback instructions; however, the other hypotheses were not supported.

Table 8

Mean (Standard Deviation) of Self-Report Carelessness per Experimental Condition

	Invisible administrator	Higher attractiveness	Lower attractiveness	Total
Basic instructions				
M (*SD*)	6.05 (.92)	6.11 (.97)	6.16 (.99)	6.11 (.95)
n	57	56	58	171
Warning instructions				
M (*SD*)	6.34 (.76)	6.32 (.94)	6.14 (1.08)	6.27 (.93)
n	53	56	51	160
Feedback instructions				
M (*SD*)	5.64 (1.23)	5.97 (1.02)	5.98 (1.05)	5.85 (1.11)
n	59	61	51	171
Total				
M (*SD*)	6.00 (1.03)	6.13 (.98)	6.09 (1.03)	
N	169	173	160	

Table 9

ANCOVA Results with Self-Report Carelessness as the Dependent Variable and Conscientiousness as a Covariate

Source	SS	df	MS	*F*	*p*	ω^2
Conscientiousness	12.19	1	12.19	17.45	.000	.0005
Instruction type	8.27	2	4.14	5.93	.003	.0004
Administrator Appearance	.97	2	.48	.69	.50	
Instructions x Appearance	.79	4	.20	.28	.89	

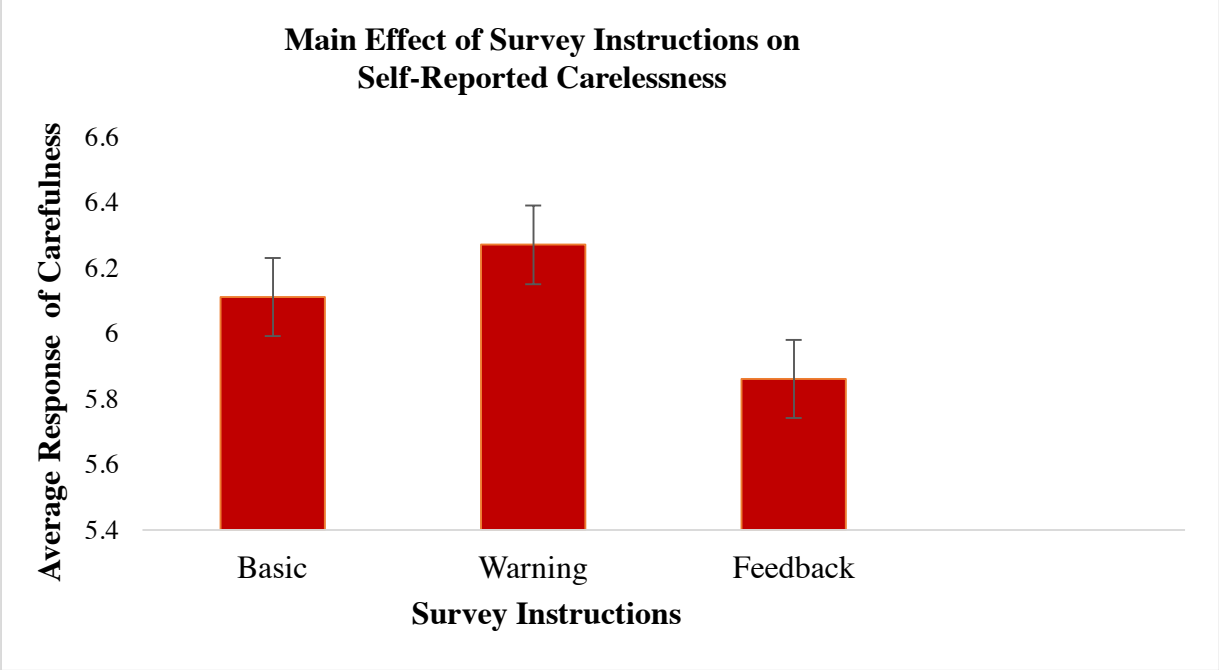

Figure 2. Self-reportedly, participants given warning instructions were significantly more careful when responding to survey items than those given feedback instructions.

Note. This item was reverse worded.

CHAPTER IV: Discussion

The purpose of the current research was to investigate the relationship between survey instructions and survey administrator appearance on measures of careless responding. It was expected that lower levels of conscientiousness and higher levels of impulsivity would relate to carelessness, respectively; and that these traits should be

controlled for when measuring careless responding behaviors in online survey taking. It was also expected that participants assigned warning instructions would provide fewer careless responses compared to those assigned basic and feedback instructions, and that participants assigned to a survey administrator higher in attractiveness would provide fewer careless responses compared to those assigned to conditions where there was no survey administrator visible, or a survey administrator lower in attractiveness. Lastly, it was expected that there would be an interaction between these two variables on the careless responding outcome measures.

Results from analyses of manipulation check items showed that respondents in the higher attractiveness conditions rated the survey administrator significantly more physically attractive than respondents in the lower attractiveness conditions. This finding is supported by past research (e.g., Coetzee, Greeff, Stevens, & Perrett, 2014) indicating that certain facial characteristics are deemed more physically attractive than others. Further, results showed that warning instructions had a significantly greater influence on respondents' level of attentiveness to survey items compared to those assigned basic instructions and feedback instructions. This finding aligns with Meade and Craig (2012) who concluded warning instructions influence attentiveness during a task.

Correlation analyses examining the relationships between conscientiousness and impulsivity on each careless responding indicator revealed that conscientiousness only significantly correlated with respondents self-reported level of carelessness, and the scales measuring impulsivity were not significantly related to any of careless responding measures used in this study. The correlation between conscientiousness and self-reported carelessness aligns with research suggesting that conscientious participants naturally

respond carefully due to their general tendency to be attentive and compliant (Meade & Pappalardo, 2013) and can attest to previous research indicating a negative relationship between conscientiousness and indices measuring insufficient-effort responding (Bowling et al., 2016).

Interpretation of Findings

The findings from this research indicated a significant interaction between survey instructions and survey administrator appearance on total response time. Posthoc analysis revealed that when given basic instructions respondents assigned to a higher attractive survey administrator, on average, took longer to respond to the survey compared to participants in both other appearance conditions. Although the main effects of each variable were not statistically significant, analysis of group means suggested that, overall, those in the higher attractiveness conditions and those given feedback instructions had the longest average response time compared to other levels of the variables. As hypothesized, these findings suggest to some extent that individuals higher in attractiveness influence observers' level of attentiveness and engagement when completing a task. Further, the finding that those given feedback instructions had longer response times aligns with Ward and Pond (2015) who found that participants given feedback instructions took longer to answer items compared to a control group.

Contrary to hypotheses, survey instructions and survey administrator appearance did not significantly influence response consistency or response patterns. Although the long string findings were not statistically significant, interesting patterns were revealed and should be noted. Respondents in the warning conditions, as well as those in the attractiveness conditions, had the lowest average long string values. Further, respondents

in the warning conditions had lower long string values at each level of the survey administrator attractiveness relative to the other experimental conditions. Though these findings were not significant, they did align with the current study's hypotheses and are partially supported by Ward and Pond (2015) who found that respondents given warning instructions had smaller maximum long string values than those given basic instructions. As mentioned previously, the extreme long string values came from respondents assigned to the basic instructions and feedback instructions. Interestingly, the respondent with the largest long string value (79, which means this person selected the same response option for all 80 items) was in the lower attractiveness/basic instructions condition.

Analysis of the self-reported measure of carelessness indicated that while controlling for conscientiousness, there was a statistically significant main effect of survey instructions on respondents' self-reported carelessness. Posthoc analysis showed that respondents given warning instructions had the highest score on this self-reported item suggesting a lesser extent of careless responding. This hypothesis was supported. As expected from previous research findings, this study showed that warning participants of a possible consequence made participants more careful when providing responses to survey items. Further, this finding aligns with results from the manipulation check item where participants given warning instructions reported a greater influence of the instructions on their level of attentiveness while completing the survey. Although the findings were not statistically significant, it should be noted that descriptive analysis showed that those assigned to the higher attractive survey administrator had the highest average score on this self-report item, also suggesting a lesser extent of carelessness.

Implications

Psychometrically speaking, statistical detection methods can never be definite indicators of carelessness. As previously mentioned, removal of respondents' data is problematic as this reduces sample sizes in a nonrandom way, artificially shapes the sample distribution, limits the generalizability of findings, and narrows the implications of the study (Ward & Pond, 2015). To limit adverse outcomes from currently used carelessness detection methods, manipulating survey design may, in part, be a viable solution to reduce the prevalence of problematic responses commonly gathered in survey research.

The findings from this research suggest that survey instructions and survey administrator's appearance do have some influence on participants' responding behavior. An implication from this research is that researchers using online survey methodology may want to include certain features in their survey design to obtain better quality data. Manipulating survey design to reduce careless responding may increase the accuracy and quality of data obtained and used in research which ultimately relates to the validity of disseminated information. If researchers opt to use survey instructions similar to those used in this study, it is important for researchers to consider the implications of threats versus following through with the instructions message. That is, if researchers continually warn participants of reducing their bonus credit or promise feedback will be provided but do not follow through, over time, these methods will likely become ineffective.

Increasing respondent engagement and attentiveness is not only important for research outcomes but has important implications for participants as well. In academic settings for instance, researchers hope to collect accurate data yet many undergraduate

participants recruited from a participant pool participate in research to obtain credits to increase their final course grade. Similarly, some respondents recruited for online surveys by an organization may choose to participate in the research because of the incentive provided (e.g., gift cards or store discounts). In these situations, participants who are extrinsically motivated may jeopardize the outcomes of the research if they are not concerned about the quality of their responses and instead satisfice to obtain the incentive. If participants are more attentive and engaged during an online survey they may get more out of the research study by, for example, reflecting on survey items and learning something new about themselves or the topic under investigation. Taken together, an important implication of the current study is that inclusion of design features that increase respondent attentiveness and engagement can create a win-win situation for researchers and respondents.

Limitations

Although several researchers have identified effective screening methods, researchers have failed to determine statistical cutoff points that would indicate definite carelessness. For instance, when measuring response patterns, there is no statistical cutoff for long string values that can indicate definite careless responders. Similarly, when measuring response time, there is no statistical cutoff point indicating a certain response time that differentiates carelessness from non-carelessness. Although studies have suggested that shortened response time indicates lack of cognitively processing items leading to carelessness (Huang et al., 2012; Ward & Pond, 2015), some interpretation is left up to the individual researcher to justify what would be considered an indication of "shortened" time span. Thus, in the current study, only comparisons between each

experimental group could be made rather than using specific cut-offs.

Another important aspect to consider is that the careless responding detection methods used in the current study may not be suitable indices for all types of survey research. For instance, use of maximum long string values as a careless responding indicator would not be appropriate for research using questionnaires pertaining to behaviors or attitudes (e.g., aggression or criminal behavior) that one would typically expect participants to repeatedly report none or very little occurrences. In such cases, consistently choosing the same response option for many or all of the questionnaire's items would not be an accurate assessment of respondent carelessness. Further, total response time may not be an appropriate careless responding indicator for some questionnaires that use survey branching. In these types of surveys, participants may be given a different number of items to respond to (based on their surveys responses to previous items) and their total response time may be affected as a result of this.

A subsequent limitation was the lack of feasibility to assess whether the gender of the survey administrator, as well as the appearance of each gender, influenced participants' responding behaviors. Including both male and female survey administrators would not only further complicate the study design by having twice the number of conditions but would also require having twice the number of participants to ensure adequate statistical power. Due to time constraints of study completion, as well as a limited number of credit hours granted by the participant pool, it would have been difficult to accomplish this task. Further, it would presumably be much more difficult to manipulate the appearance of a male researcher to appear more or less attractive with the use or misuse of makeup. Although failure to use both male and female researchers in the

current study may limit our understanding of whether the gender and level of attractiveness influence students' responding in online surveys, research examining the relationship between a teacher's level of attractiveness and teaching evaluations as rated by students, have shown that the impact of teacher beauty on student engagement was significant for both male and female faculty (Hamermesh & Parker, 2005).

Future Directions

Future research should examine survey administrator appearance and survey instructions using samples from different populations. Researchers interested in the influence of physical appearance on observers' behavior can compare survey administrators of varying genders and ages on different samples of respondents. It would be interesting to assess whether the gender and/or age of the survey administrator is more effective in increasing participant engagement and attentiveness and whether these characteristics are better suited for certain types of participants or topics of online surveys.

A common finding within this line of research are that warning instructions are effective at increasing attentiveness and reducing carelessness. In this study, undergraduate students participated in exchange for bonus credit added to their final grade so it is likely that the warning instructions were influential on responding behavior due to respondents' belief that they may have received reduced bonus credit for carelessness. It may be the case that feedback instructions are more influential for survey research recruiting samples that are not offered any incentive for participation. As noted by Gosling, Vazire, Srivastava, and John (2004), feedback appeals to individuals' desire for self-insight and participants are motivated to answer honestly to receive accurate

feedback about themselves and/or their performance. In the current study, the feedback survey instructions may have been more effective in increasing engagement if participants were given feedback about something important or interesting to them, for instance their personality profile. Feedback instructions appear to be an underexplored area of research and should be further examined as this type of instruction is a nonaversive way for participants to potentially provide better quality data.

References

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron, 32*, 537-551.

Akbulut, Y. (2015). Predictors of inconsistent responding in web surveys. *Internet Research, 25*(1), 131-147.

Agnew, R. (1984). The effect of appearance on personality and behavior: Are the beautiful really good? *Youth and Society, 15*(3), 285-303.

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment and cooperation: A meta-analysis. *Psychological Bulletin, 137*(4), 593-615.

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education, 53*(2), 182–200.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Medicine and Health, 10(3),* 295-307.

Behrend, T., & Foster-Thompson, L. (2011). Similarity effects in online training: Effects with computerized trainer agents. *Computers in Human Behavior, 27*, 1201–1206.

Berinsky, A. J., Margolis, M. F., Sances, M. W. (2013). Separating the shrinkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Politics Science, 58*(3), 739-753.

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology.* Advance Online Publication.

Burnett, P. C., & Fanshawe, J. P. (1997). Measuring school-related stressors in adolescents. *Journal of Youth and Adolescence, 26*(4), 415-428

Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment, 83*(1), 29-45.

Chaikin, A. L., Gillen, B., Derlega, V. J., Heinen, J. R., & Wilson, M. (1978). Students' reactions to teachers' physical attractiveness and nonverbal behavior: Two exploratory studies. *Psychology in the Schools*, *15*(4), 588-595.

Chambel, M. J., & Curral, L. (2005). Stress in academic life: Work characteristics as predictors of student well-being and performance. *Applied Psychology, 54(*1), 135-147.

Clifford, S., & Jerit, J. (2015). Do attempts to improve respondent attention increase social desirability bias? *Public Opinion Quarterly, 79*(3), 790-802.

Coetzee, V., Greeff, J. M., Stephen, I. D., & Perrett, D. I. (2014). Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender. Plos One 9: e99629.

Colledge, E., & Blair, R. J. R. (2001). The relationship in children between the inattention and impulsivity components of attention deficit and hyperactivity disorder and psychopathic tendencies. *Personality and Individual Differences, 30*(7), 1175-1187.

Curran, P. G., Kotrba, L., & Denison, D. (2010, April). *Careless responding in surveys: applying traditional techniques to organizational settings*. Paper presented at

the 25th annual conference of Society for Industrial and Organizational Psychology, Atlanta, GA.

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36*(2), 171-181.

Diekhof, E. K., Nerenberg, L., Falkai, P., Dechant, P., Baudewig, J., & Gruber, O. (2011). Impulsive personality and the ability t resist immediate reward: An fMRI study examining inter-individual differences in the neural mechanisms underlying self-control. *Human Brain Mapping, 33*(12) 2768-2784.

Dion, K. K., & Berscheid, E. (1974). Physical attractiveness and peer perception among children. *Sociometry, 37*(1), 1-12.

Dion, K., Berscheid, E., & Walster, E. (1975). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285-290.

Gehlbach, H., & Barge, S. (2012). Anchoring and adjusting in questionnaire responses. *Basic and Applied Social Psychology, 34*(5), 417–433.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34.

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 Psychology, 59*(2), 93-104.

Gurung, R. A. R., & Vespia, K. (2007). Looking good, teaching well? Linking liking, looks and learning. *Teaching of Psychology, 34*(1), 5-10.

Hauser, D. J., Sunderrajan, A., Natarajan, M., & Schwarz, N. (2016). Prior exposure to

instructional manipulation checks does not attenuate survey context effects driven by satisficing or gricean norms. *Methods, Data, Analyses, 10*(2), 195-220.

Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *Public Opinion Quarterly, 3*(71), 1-25)

Huang, J., Bowling, N., Liu, M., & Li, Y. (2014). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business Psychology, 30*, 299-311.

Huang, J., Curran, P., Keeney, J., Poposki, E., & DeShon, R. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114.

Jackson, D., Singleton-Jackson, J., Frey, M., & Mclellan, C. (2013). Understanding and defining academic entitlement. Poster presented at the annual Conference on Teaching and Learning, Rochester, MI.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web- based personality inventories. *Journal of Research in Personality, 39*(1), 103–129.

Kohn, J. P., Frazer, G. H. (1986). An academic stress scale: Identification and rate importance of academic stressors. *Psychological Reports, 59*, 415-426.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2)*,* 254–284.

Krosnick, J. A. (1991). Response strategies for coping with cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236.

Krosnick, Jon A. 2000. "The Threat of Satisficing in Surveys: The Shortcuts Respondents Take in Answering Questions." *Survey Methods Newsletter* 20(1): 4–8.

Kubanek, J., Synder, L. H., & Abrams, R. A. (2015). Reward and punishment act as distinct factors in guiding behavior. *Cognition, 139*, 154-167.

Langlois, J. H., Ritter, J. M., Roggman, L. A., & Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology, 27*(1), 79-84.

Lucas, R. E., & Baird, B. M. (2005). *Global self-assessment. In M. Eid & E. Diener* (Eds.), Handbook of multimethod measurement in psychology (pp. 29–42). Washington, DC: American Psychological Association.

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood: Physical attractiveness and accuracy in first impressions of personality. *Psychological Science, 21*(12), 1777-1782.

Lucker, G. W., Beane, W. E., & Helmreich, R. L. (1981). The strength of the halo effect in physical attractiveness research. *The Journal of Psychology, 107*, 69-75.

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Personality, 48*, 61-83.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as source of error variance in applied assessment. *Psychological Bulletin, 136*(3), 450–470.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455.

Meade, A. W., & Pappalardo, G. (2013). *Predicting careless responses as attrition in*

survey data with personality. Paper presented at the 28th Annual Meeting of Society for Industrial and Organizational Psychology, Houston, TX.

Miller, J., & Baker-Prewitt, J. (2009). Beyond 'trapping' the undesirable panelist: The use of red herrings to reduce satisficing. Paper presented at the CASRO Quality Conference.

Northcraft, G. B., Schmidt, A. M., & Ashford, S. J. (2011). Feedback and the rationing of time and effort among competing tasks. Journal of Applied Psychology, 96(5), 1076–1086.

Patton, J. H., & Stanford, M. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768-774.

Perloff, R. M. (2003). *The dynamics of persuasion: Communication and attitudes in the 21st century*. Mahwah, N.J: Lawrence Erlbaum Associates.

Reinhardt, J (2011). Conceptualizing academic entitlement: What are we measuring? *Unpublished Master's Thesis, University of Windsor.*

Reise, S. P., Moore, T. M., Sabb, F. W., Brown, A. K., & London, E. D. (2014). The barratt impulsiveness scale-11: Reassessment of its structure in a community sample. *Psychological Assessment, 25*(2), 631-642.

Richardson, M., & Abraham, C. (2009). Conscientiousness and achievement motivation predict performance. *European Journal of Personality, 23*(7), 589-605.

Riggle, E. D. B., Rostosky, S. S., & Reedy, C. S. (2005). Online surveys for BGLT research. *Journal of Homosexuality, 49*(2), 1-21.

Riniolo, T. C., Johnson, K. C., Sherman, T. R., & Misso, J. A. (2006). Hot or not: Do professors perceived as physically attractive rreceive higher student evaluations?

The Journal of General Psychology, 133(1), 19-35.

Robinson-Cimpian, J. (2014). Inaccurate estimation of disparities due to mischievous responders: Several suggestions to assess conclusions. *Educational Researcher, 43*(4), 171–185.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. American Psychologist, 54, 93–105.

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridical judgment. *Journal of Personality and Social Psychology, 31*, 410-440.

Simon, H. A. (1957). *Models of man.* New York: Wiley.

Skinner, B. F. (1938). *The behavior of organisms. An experimental analysis.* New York: Appleton-Century.

Soto C, & John, O. (2009). Ten facet scales for the big five inventory: Convergence with NEO-PI-R facets, self-peer agreement, and discriminant validity. Journal of Research in Personality, 43, 84-90.

Sui. J., & Liu, C. (2009). Can beauty be ignored? Effects of facial attractiveness on cover attention. *Psychonomic Bulletin & Review, 16*(2), 276-281.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Vanette, D. Using attention checks in your surveys mat harm data quality. (June 2017). Qualtrics Blog retrieved from: https://www.qualtrics.com/blog/using-attention-checks-in-your-surveys-may-harm-data-quality/

Ward, P., Clark, T., Zabriskie, R., & Morris, T. (2014). Paper/pencil versus online data

collection: An exploratory study. *Journal of Leisure Research, 46*(1), 84-105.

Ward, M. K., & Pond, S. (2015). Using virtual presence and survey instructions to minimize careless responding on internet-based surveys. *Computers in Human Behavior, 48*, 554-568.

Weijters, B., De Beuckelaer, A., & Baumgartner, H. (2014). Discriminant validity where there should be none: Positioning same-scale items in separated blocks of a questionnaire. *Applied Psychological Measurement, 38*, 450–463.

Westfall, R. S. (2015). Effects of instructor attractiveness on classroom learning. UNLV Theses, Dissertations, Professional Papers, and Capstones. 2446.

Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3), 1.

Zwarun, L., & Hall, A. (2014). What's going on? Age, distraction, and multitasking during online survey taking. *Computers in Human Behavior, 41*, 236–244.

Appendix A
The Big Five Inventory (BFI)

Here are several characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number in the box next to each statement to indicate the extent to which you agree or disagree with that statement.

Disagree strongly	**Disagree a little**	**Neither agree nor disagree**	**Agree a little**	**Agree Strongly**
1	2	3	4	5

I see myself as someone who...

_____ 1. Is talkative

_____ 2. Tends to find fault with others

_____ 3. Does a thorough job

_____ 4. Is depressed, blue

_____ 5. Is original, comes up with new ideas

_____ 6. Is reserved

_____ 7. Is helpful and unselfish with others

_____ 8. Can be somewhat careless

_____ 9. Is relaxed, handles stress well

_____ 10. Is curious about many different things

_____ 11. Is full of energy

_____ 12. Starts quarrels with others

_____ 13. Is a reliable worker

_____ 14. Can be tense

_____ 15. Is ingenious, a deep thinker

_____ 16. Generates a lot of enthusiasm

_____ 17. Has a forgiving nature

_____ 18. Tends to be disorganized

_____ 19. Worries a lot

_____ 20. Has an active imagination

_____ 21. Tends to be quiet

_____ 22. Is generally trusting

_____ 23. Tends to be lazy
_____ 24. Is emotionally stable, not easily upset
_____ 25. Is inventive
_____ 26. Has an assertive personality
_____ 27. Can be cold and aloof
_____ 28. Perseveres until the task is finished
_____ 29. Can be moody
_____ 30. Values artistic, aesthetic experiences
_____ 31. Is sometimes shy, inhibited
_____ 32. Is considerate and kind to almost everyone
_____ 33. Does things efficiently
_____ 34. Remains calm in tense situations
_____ 35. Prefers work that is routine
_____ 36. Is outgoing, sociable
_____ 37. Is sometimes rude to others
_____ 38. Makes plans and follows through with them
_____ 39. Gets nervous easily
_____ 40. Likes to reflect, play with ideas
_____ 41. Has few artistic interests
_____ 42. Likes to cooperate with others
_____ 43. Is easily distracted
_____ 44. Is sophisticated in art, music, or literature

Appendix B
Barratt's Impulsiveness Scale

People differ in the ways they act and think in different situations. This is a test to measure some of the ways in which you act and think. Please select the option best represents your answer. Do not spend too much time on any statement. Answer quickly and honestly.

Statement	Rarely/ Never	On Occasion	Often	Almost Always/ Always
1. I plan tasks carefully.	1	2	3	4
2. I do things without thinking.	1	2	3	4
3. I make-up my mind quickly.	1	2	3	4
4. I am happy-go-lucky.	1	2	3	4
5. I don't "pay attention."	1	2	3	4
6. I have "racing" thoughts.	1	2	3	4
7. I plan trips well ahead of time.	1	2	3	4
8. I am self controlled.	1	2	3	4
9. I concentrate easily.	1	2	3	4
10. I save regularly.	1	2	3	4
11. I "squirm" at plays or lectures.	1	2	3	4
12. I am a careful thinker.	1	2	3	4
13. I plan for job security.	1	2	3	4
14. I say things without thinking.	1	2	3	4
15. I like to think about complex problems.	1	2	3	4
16. I change jobs.	1	2	3	4
17. I act "on impulse".	1	2	3	4
18. I get easily bored when solving thought problems.	1	2	3	4
19. I act on the spur of the moment.	1	2	3	4
20. I am a steady thinker.	1	2	3	4
21. I change residences.	1	2	3	4
22. I buy things on impulse.	1	2	3	4
23. I can only think about one thing at a time.	1	2	3	4
24. I change hobbies.	1	2	3	4
25. I spend or charge more than I earn.	1	2	3	4
26. I often have extraneous thoughts when thinking.	1	2	3	4
27. I am more interested in the present than the future.	1	2	3	4
28. I am restless at the theater or lectures.	1	2	3	4
29. I like puzzles.	1	2	3	4
30. I am future oriented.	1	2	3	4

Appendix C
Academic Stress Scale

Ranging from 0 to 100, please indicate how stressful each of the following academic events are to you. If the event is considered more stressful than taking an exam, rate the item between 51 and 100. If the event is considered less stressful to you than taking an exam, please rate it between 0 and 49. If the event is considered as stressful as taking an exam, please rate it 50.

	Event	**Rating (0-100)**
1	Taking exams	
2	Being unprepared to respond to questions	
3	Attending boring classes	
4	Taking an announced quiz	
5	Receiving final grades	
6	Taking a pop quiz	
7	Writing a term paper	
8	Taking irrelevant classes toward major	
9	Taking classes with open discussions	
10	Having excessive homework	
11	Evaluating classmates' work	
12	Taking notes in class	
13	Forgetting to complete an assignment	
14	Handing in an incomplete assignment	
15	Speaking in class	
16	Arriving late for class	
17	Being dismissed late from class	
18	Being in a noisy classroom	
19	Being in a hot classroom	
20	Being in a cold classroom	
21	Being in a crowded classroom	
22	Being in a poorly lit classroom	
23	Learning new skills	
24	Missing classes	
25	Buying textbooks	
26	Studying for an exam	
27	Non-native language lectures	
28	Reading the wrong material	
29	Being in fast-paced lectures	
30	Forgetting pen/pencils	
31	Being given an unclear assignment	
32	Being given an unclear course objective	
33	Giving incorrect answers in class	
34	Attending the wrong class	
35	Waiting for test grades	

Appendix D
Academic Well-Being Scale

Please indicate the extent to which you during the preceding month your academic work made you feel the following way:

	Item	**Never**						**Always**
1	Tense	1	2	3	4	5	6	7
2	Anxious	1	2	3	4	5	6	7
3	Worried	1	2	3	4	5	6	7
4	Calm	1	2	3	4	5	6	7
5	Comfortable	1	2	3	4	5	6	7
6	Relaxed	1	2	3	4	5	6	7
7	Depressed	1	2	3	4	5	6	7
8	Gloomy	1	2	3	4	5	6	7
9	Miserable	1	2	3	4	5	6	7
10	Cheerful	1	2	3	4	5	6	7

Appendix E
Psychological Entitlement Questionnaire

This questionnaire is used to measure some of the ways in which you act and think. Please select the option that best represents your answer.

Statement	**Strongly Disagree**						**Strongly Agree**
1. I honestly feel I'm just more deserving than others.	1	2	3	4	5	6	7
2. Great things should come to me.	1	2	3	4	5	6	7
3. If I were on the Titanic, I would deserve to be on the *first* lifeboat!	1	2	3	4	5	6	7
4. I demand the best because I'm worth it.	1	2	3	4	5	6	7
5. I do not necessarily deserve special treatment.	1	2	3	4	5	6	7
6. I deserve more things in my life.	1	2	3	4	5	6	7
7. People like me deserve an extra break now and then.	1	2	3	4	5	6	7
8. Things should go my way.	1	2	3	4	5	6	7
9. I feel entitled to more of everything.	1	2	3	4	5	6	7

Appendix F
Academic Entitlement Questionnaire

Please indicate the extent to which you identify with the following statements.

	Question	Strongly Disagree						Strongly Agree
1	Great academic success should just come to me.	1	2	3	4	5	6	7
2	I do not necessarily deserve special treatment from my professors.	1	2	3	4	5	6	7
3	I deserve more praise from my professors.	1	2	3	4	5	6	7
4	If a professor were only allowed to give one “A” in a course, it should be given to me.	1	2	3	4	5	6	7
5	I honestly feel I am more deserving than other students.	1	2	3	4	5	6	7
6	I demand the best grades because I deserve them.	1	2	3	4	5	6	7
7	I deserve more A’s.	1	2	3	4	5	6	7
8	My effort in a course should be considered in the final grade.	1	2	3	4	5	6	7
9	I deserve a passing grade for attending all lectures in a course.	1	2	3	4	5	6	7
10	I should never fail an assignment I put effort into.	1	2	3	4	5	6	7
11	If I have attended most classes for a course, I deserve a good grade.	1	2	3	4	5	6	7
12	Professors should not round up my grade based on effort.	1	2	3	4	5	6	7
13	If I have completed most of the reading for a class, I deserve a good grade.	1	2	3	4	5	6	7
14	When assigning my course grade, my professor should consider how hard I have tried.	1	2	3	4	5	6	7
15	It is only the quality of my work that matters when assigning grades.	1	2	3	4	5	6	7
16	Professors should bend the rules for me.	1	2	3	4	5	6	7
17	If I do not complete my work on time, I do not deserve to be able to hand it in late.	1	2	3	4	5	6	7
18	Professors should not put material on a test that students have trouble understanding.	1	2	3	4	5	6	7
19	My test date should be moved if I am not prepared.	1	2	3	4	5	6	7
20	If I am unable to complete my assignment on time I should still be able to hand it in by the last day of class.	1	2	3	4	5	6	7
21	I should not have to think too hard to learn the material for a class.	1	2	3	4	5	6	7
22	I should not be given special treatment to help me perform better in a class.	1	2	3	4	5	6	7

23	A professor should modify course requirements to help me achieve a better grade.	1	2	3	4	5	6	7
24	I am not motivated to put effort into group work, because another group member will end up doing the work.	1	2	3	4	5	6	7
25	If I miss a test I should not have to explain to the professor why.	1	2	3	4	5	6	7
26	If I do poorly in a course, the fault lies with my professor.	1	2	3	4	5	6	7
27	It is my responsibility to seek out the resources to succeed in university.	1	2	3	4	5	6	7
28	For group assignments, it is acceptable to take a back seat and let others do most of the work.	1	2	3	4	5	6	7
29	It is acceptable to lie to a professor if it helps me avoid failing an assignment.	1	2	3	4	5	6	7
30	I should receive the same grade as the other group members regardless of my level of effort.	1	2	3	4	5	6	7
31	If I miss class, it is my responsibility to catch up on the material I missed.	1	2	3	4	5	6	7
32	Professors work for students.	1	2	3	4	5	6	7
33	I am a customer of this university.	1	2	3	4	5	6	7
34	My professors are not obligated to hold special test preparation sessions.	1	2	3	4	5	6	7
35	I should be responsible for knowing assigned reading material even if it is not discussed in class.	1	2	3	4	5	6	7
36	Professors are just employees who get money for teaching.	1	2	3	4	5	6	7
37	Information on exams should be entirely based on material taught to me in lecture.	1	2	3	4	5	6	7
38	I deserve to be entertained by my professors' lectures.	1	2	3	4	5	6	7
39	I deserve to have more input in how my classes are taught	1	2	3	4	5	6	7
40	I would think poorly of a professor who did not respond the same day to an e-mail I sent.	1	2	3	4	5	6	7
41	A professor should be willing to meet with me at a time that works best for me, even if inconvenient for the professor.	1	2	3	4	5	6	7
42	When my personal plans conflict with an exam the professor should not let me take the exam at a different time.	1	2	3	4	5	6	7
43	A professor should not tolerate students receiving telephone calls in class.	1	2	3	4	5	6	7
44	Professors should respond to e-mails within 30 minutes.	1	2	3	4	5	6	7

45	A professor should let me arrange to turn in an assignment late if the due date interferes with my personal plans.	1	2	3	4	5	6	7
46	I should be able to call my professor at home if I need help.	1	2	3	4	5	6	7
47	A professor should be willing to provide his or her course notes to me if I ask for them.	1	2	3	4	5	6	7
48	I would think poorly of a professor who did not respond quickly to a voice mail I left him or her.	1	2	3	4	5	6	7
49	There is nothing wrong with arguing to get more points on a test.	1	2	3	4	5	6	7
50	It is acceptable to demand higher grades from my professors.	1	2	3	4	5	6	7
51	Asking for extra points on assignments is an acceptable strategy to improve my grades.							
52	The grades I receive accurately reflect what I have learned.	1	2	3	4	5	6	7
53	It is acceptable to confront a professor to argue about my grade.	1	2	3	4	5	6	7
54	No tactic is too extreme when arguing for an improved grade.	1	2	3	4	5	6	7
55	Professors just make grades up, so it is not a problem to argue for a higher grade.	1	2	3	4	5	6	7
56	I always deserve a higher grade than I am given, making it necessary to argue for extra points.	1	2	3	4	5	6	7
57	I should earn my grades not argue for them.	1	2	3	4	5	6	7
58	Students should complain to the Dean or higher level of authority to get the grade they want.	1	2	3	4	5	6	7
59	Professors should raise my grade to prevent me from losing a scholarship.	1	2	3	4	5	6	7
60	Professors should raise my grade to prevent me from being placed on academic probation.	1	2	3	4	5	6	7
61	A professor should never raise grades once they are assigned, even if he or she made an error.	1	2	3	4	5	6	7

Appendix G
Manipulation Check Items and Self-Reported Carelessness

Please rate the survey administrator's physical appearance on a scale from 1 (*not at all physically attractive*) to 10 (*very physically attractive*).

Not at all physically attractive									Very physically attractive
1	2	3	4	5	6	7	8	9	10

Would you generally consider the survey administrator to be lower in physical attractiveness, average, or higher in physical attractiveness?"

Lower in physical attractiveness	Average	Higher in physical attractiveness
1	2	3

To what extent did the survey instructions that you received influence your level of carefulness when responding to the survey items?

Not at all						Very Much
1	2	3	4	5	6	7

To what extent do you think your responses reflects are accurate and are of sufficient quality for researchers to use?

Very poor quality						Very good quality
1	2	3	4	5	6	7

Appendix H
Demographic Questionnaire

Please answer the following questions to help us classify your responses.

1. What is your age? _____

2. What is your Gender?

- Male
- Female
- Transgender

3. Are you a first year student at the University of Windsor – that is, did you begin taking classes here in the Fall of 2016?

- Yes
- No

4. Have you taken university courses prior to attending the University of Windsor?

- Yes
- No

If yes, when? (e.g., 2011 to 2012) _____

5. What is your major area of study? _____

6. Ethnicity: _____

7. Student Status
Are you a:

- Canadian Student
- American Student
- International Student

8. Current year of study:

- 1
- 2
- 3
- 4 or 4+

9. In which country were you born?

10. In which country was your mother born?

11. In which country was your father born?

12. What is the approximate yearly income for your family household?

13. What is your cumulative GPA?

14. With respect to your performance in a typical class, would you say that you typically perform…

- In the top 10%
- In the top 25%, but not the top 10%
- In the top half, but not the top 25%
- In the bottom half

15. On the last exam you took, would you say that your performance was…

- In the top 10%
- In the top 25%, but not the top 10%
- In the top half, but not the top 25%
- In the bottom half

16. What was the score you received on the last assignment for which you received feedback?

- 90% or higher
- 80-89%
- 70-79%
- 60-69%
- 50-59%
- Below 50%

17. When you compare your grades to those of your friends and classmates, are your grades typically…

- Much higher than others'
- Usually a little bit higher than others'

- About the same as others'
- Usually a little bit lower than others'
- Much lower than others'

18. Approximately how many hours per week do you spend studying <u>alone</u> for your courses?

- Fewer than 2 hours per week
- 2 to 5 hours per week
- 6 to 10 hours per week
- 11 to 15 hours per week
- More than 15 hours per week

19. Approximately how many hours per week do you spend studying with a friend or with a group of people for your courses?

- Fewer than 2 hours per week
- 2 to 5 hours per week
- 6 to 10 hours per week
- 11 to 15 hours per week
- More than 15 hours per week

Vita Auctoris

Name:	Carolyn Marie Rauti
Birth Place:	Windsor, Ontario, Canada
Year of Birth:	1989
Education:	University of Western Ontario, B. A London, Ontario, Canada 2011 - 2015
	University of Windsor, M. A Windsor, Ontario, Canada 2015 – 2017
Funding:	Social Sciences and Humanities Research Council