TIRED OF SURVEY FATIGUE?
INSUFFICIENT EFFORT RESPONDING DUE TO SURVEY FATIGUE

by

Hung Loan T. Nguyen

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of Master of Arts in
Industrial Organizational Psychology

Middle Tennessee State University

May 2017

Thesis Committee:

Dr. Michael B. Hein, Chair

Dr. Mark C. Frame, Member

Dr. Judith Van Hein, Critical Reader

I dedicate this research to my loving family. Thanks for everything.

ACKNOWLEDGMENTS

I would like to thank my family for providing guidance and supporting me throughout this experience. I would also like to thank my amazing cohort for always pushing me. I think very highly of all of you and would consider all of you my close friends. Finally, I would like to thank Dr. Michael Hein, Dr. Mark Frame, and Dr. Judith Van Hein for providing me great advice and constructive feedback throughout this experience.

ABSTRACT

Insufficient effort responding (IER) has been found to be prevalent in a variety of settings and have also been found to have major impacts on survey data quality. One way to prevent IER from occurring is by engaging in preventative measures to reduce different types of IER. One possible preventative measure that can be done to prevent IER is reducing the length of long surveys. To support this notion, this study looked at the effects of survey length and data quality.

Results indicated that long surveys can negatively impact data quality. Specifically, factor structures of measures placed at the end of a long survey may be negatively impacted. Additionally, the reliability of measures decreased when placed later in a long survey. Lastly, results indicated that total time to survey completion was positively related to answering quality assurance items designed to detect IER.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

I. LITERATURE REVIEW

Overview

Surveys are being conducted more frequently and in a variety of settings. Information from surveys are frequently being used to make major decisions so it is important for them to be valid (Credé, 2010). Research that uses survey data assumes that responders are providing quality data; however, it has been found that responders do not always provide the best data possible (Huang et al., 2012; Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012; Nichols, Greene, & Schmolck, 1989). Often survey participants engage in bad responding; the three most researched types of bad responding are socially desirable (faking), acquiescence, and careless and/or random responding (Furnham, Hyde, & Trickey, 2015). Furthermore, Zieglar, MacCann and Roberts (2012) stated that the majority of research on bad responding has focused on faking and there has been less of a focus on careless and/or random responding.

Nichols, Greene and Schmolck (1989) categorized bad responding into two categories: content responsive faking (e.g., social desirability and impression management) and content nonresponsivity. Content responsive faking is defined as responding in a way that "over endorses items, complaints, negative and positive psychological attributes" (pg. 241). Unlike faking, content nonresponsivity is defined as responding in a way that ignores the contents of the item "regardless of motivation" and can be due to "unwillingness to comply with survey procedures and even linguistic incompetence" (pg. 240). Often in research, content faking and nonresponsivity are lumped together as similar phenomenon but there is evidence to indicate that they are separate constructs. In fact, Maniaci and Rogge (2014) found that social desirability and

impression management were actually negatively correlated with inattentiveness and noncompliance which indicates they should be treated as separate constructs. With that being said, there has been growing interest on the topic of content nonresponsivity.

As a relatively young area, research done on content nonresponsivity has been inconsistent with labels. The nomenclature used to refer to content nonresponsivity has ranged from "careless and/or inattentive" to "random" responding, with the most common being random responding (Huang et al., 2012). This is a problem in that content nonresponsivity can result in different response patterns and are not necessarily random (Meade & Craig, 2012) and it would be inappropriate to label all content nonresponsivity as random responding. For consistency and accuracy, building on Nichols, Green, and Schmolck's work and similar work in the area (Beach, 1989; Curran, Kotraba, & Denison, 2010; Johnson, 2005), Huang, Curran, Keeney, Poposki, and DeShon (2012) coined the term Insufficient Effort Responding (IER) to refer to all types of content nonresponsivity. IER is defined as a "specific response set in which the responder responds to survey measures with low or little motivation to comply with survey instructions, interpret item contents, or to provide accurate responses" (Huang et al., 2012, pg. 100). This term is broader and captures both random and inattentive/careless responding. Creating the term IER establishes a common accepted nomenclature and allows researchers to focus on how to deal with IER—dealing with IER is another area that is debated.

IER as a whole has been found to be prevalent in a variety of settings and have major implications on survey data quality. However, most research on IER has looked at IER as a nuisance and has focused on finding ways to detect IER in order to remove bad

data from the dataset (e.g., Huang, Bowling, & Liu, 2014; Maniaci & Rogge, 2014; Meade & Craig, 2012). The problem with this methodology is that IER has already occurred. Piedmont, McCrae, Riemann, and Angleitner (2000) discredited the use of the detection methods and proposed that focus should shift from detecting IER to improving the quality of assessments, motivating participants, and making sure participants understand instructions. One possible preventative measure that can be taken to decrease the risk of IER is decreasing survey length in long surveys.

There is little research that investigates what constitutes a "long survey". Most of the research on long surveys focus on the effects of long surveys on response rates but little has been done to look at the effects of survey length at the measure and item level. Dillman (1978) recommended surveys be limited to 10 pages or about 125 items or less to increase response rates. Galesic and Bsonjak (2009) operationally defined a long survey as a survey that takes an average of 30-45 minutes to complete but Herzog and Bachman (1981) defined a long survey as a survey that takes an average of 2-hours to complete. Due to the conflicting definition of long surveys, this study will combine these three definitions. A long survey, for the purposes of this study, was defined as a survey that has more than 125 items and takes at least an average of 1-hour to complete.

Survey length is beneficial in increasing reliability of measures. However, there may be a point of diminishing returns. Several studies have suggested that IER may occur more frequently towards the end of long surveys than at the beginning (Baer, Ballenger, Berry, & Wetter, 1997; Berry et al., 1992; Meade & Craig, 2012). According to Meade and Craig (2012), participants may become increasingly inattentive as they complete longer surveys. Krosnick (1991) proposed that as survey participants complete surveys,

their cognitive resources are depleted and will start to unconsciously select response strategies that limit cognitive effort, which may result in lower response quality. This was coined “satisficing” but for the purpose of this manuscript will be called “survey fatigue”. A possible solution to prevent survey fatigue from occurring is decreasing the length of long surveys but this may in turn decrease reliability (DeVellis, 1991). Due to constraints for both the preventative and detection approaches, it may be more fruitful to combine the approaches.

The current manuscript gives merit to both the preventative and the detection approaches. Combining the approaches is ideal in that if researchers can detect and distinguish different distinct forms and properties of IER, then action can be taken in creating and changing characteristics of assessments to reduce or prevent those specific IERs from occurring in the first place. As the first step in combining the preventative and detection approach, this study attempted to distinguish IER due to survey fatigue as a distinct and unique type of IER.

Camus (2015) found that IER is stable across time but not across task. This means that IER may be caused in part by specific survey characteristics, which provides support to the idea that IER may be accounted for when creating a survey. If IER due to survey fatigue is shown to be distinct from other forms of IER then researchers and survey creators can design surveys to be shorter to account for survey fatigue. Another possibility is researchers may want to acknowledge item positioning and place more important/complex items or measures towards the beginning of a long survey to compensate for less valid responses at the end of longer surveys. It also may mean that IER is a bigger issue with long surveys conducted online and on micro task websites like

MTURK which have been found to already have a higher prevalence of IER than traditional paper and pencil surveys (Fleischer, Meade, & Huang, 2015; Johnson, 2005; Meade, 2012). Identifying IER due to survey fatigue may have merit but the notion behind distinct IER is contingent on whether or not there is evidence to indicate different forms of IER exist.

The first step in distinguishing different forms of IER is determining if there is evidence for unique and distinct forms of IER. One way this can be done is by comparing the prevalence rates and psychometric consequences that have been reported. In general, the reported prevalence and impact of IER have varied.

Prevalence and Impact of IER. Liu, Bowling, Huang and Kent (2013) conducted a survey to investigate the perceptions of the impact on data quality and the prevalence of IER from 253 Society for Industrial and Organizational Psychology (SIOP) members and found that these specific members perceived IER as a minor or moderate issue and do little to address the threat. This suggests that many SIOP members perceive IER as a mere nuisance with little impact on data quality; this greatly contrasts with the growing body of evidence that says otherwise. The rate of IER indicated by various studies indicates that IER is a bigger problem than what many SIOP members believe.

The rate of IER has varied drastically between different studies but what they all indicate is that IER is prevalent in different settings. Generally, IER has been found to occur between 23-47% of all responders (Fleischer, Mead, & Huang, 2015; Meade & Craig, 2011; Meade & Pappalardo, 2013; Oppenheimer, Meyvis, & Davidenko, 2009). One reason why there is such a large range in the reported prevalence of IER is that researchers often operationally define IER differently, leading to different detection

methods that have different efficacies in detecting IER, refer to the Detection of IER section below for further information on this matter. In contrast, the prevalence of extreme IER, response data that is meaningless or nearly, has been found to be between 2-12% (Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012; Meade & Papparlardo, 2013). This means that occasional IER occurs frequently but only a small portion of responders actually engage in extreme IER. Even though there is only a small percentage of responders who engage in extreme IER, general IER has been found to have a more profound impact on data quality than extreme IER. Using archival data, Holden, Wheeler, and Marjanovic (2012) found that IER moderated the validity of responses on the NEO PI-R only when it was at a non-extreme level suggesting that general IER may have a more negative impact on response data than extreme IER. One specific impact IER can have on survey data is by altering the factor structures of measures.

IER may change the factor structures of measures. When negatively worded items (i.e., reverse worded items) are present, both Schmitt and Stults (1986) and Woods (2006) have found that a rate of 10% IER will lead to an additional distinct 'negatively worded factor' in unidimensional scales. This means that negatively worded items are not measuring what they are intended to measure and factor onto separate constructs from their regularly worded counterparts. In organizational settings, Merrit (2012) found that Allen and Meyers' (1990) unidimensional Affective Commitment Scale has a two-factor solution when there are negatively worded items and IER present and also found that factor structures were effected even when negatively worded items were not present (albeit to a lesser degree). Another organizational example, with potential major

significance, is a study conducted by Kam and Meyer (2015). Herzberg and his colleagues (1959) claimed that job satisfaction and job dissatisfaction are distinct constructs rather than polar opposites which was supported by Credé et al. (2009). Kam and Meyer (2015) found conflicting evidence when replicating Credé et al.'s study. In their study, they found that job satisfaction and job dissatisfaction were polar opposites of the same construct when controlling for IER. In addition, the researchers proposed that random responding may not be as harmful as systematic responding (e.g., selecting 'strongly agree' for all items) because random responding only leads to weaker associations within variables but does not change the factor structures of constructs, whereas systematic response patterns (non-random) can change factor structures when there are negatively worded items and 10% careless responding present. This is of interest because there is evidence to indicate that responses towards the end of long surveys have a tendency to be towards central response options and typically are not random (Herzog & Bachman, 1981). These studies provide credence to the idea that IER can impact data quality of surveys by altering factor structures of measures and further distinguishes random and nonrandom IER. Another effect IER has on data quality is decreasing statistical power.

IER has been found to increase Type II error rates. According to the classical test theory, IER is random error and is a source of "noise" and increases chance of Type II error. This noise is supposed to attenuate the magnitude of associations between variables. Oppenheimer, Meyvis, and Davidenko (2009) found support for this by inserting their instructional manipulation checks into surveys with two robust paradigms from the judgement and decision literature in the form of Thaler's (1954) Beer Pricing

Task and Football Sunk Cost Question. They found that when data from participants flagged for IER were removed from the dataset, their findings were significant or amplified, but when they weren't accounted for that their results were either insignificant or attenuated. Furthermore, Maniaci and Rogge (2014) also found that higher proportions of IER reduced power in their simulated study by as much as 36% and reduced effect sizes significantly. They also found that screening for IER resulted in increased power by as much as 24%. These studies indicate that IER can be a source of noise and thus increase Type II error rates. Although most research has viewed IER as a source of noise, it has also been observed to increase the chance of Type I error rates.

IER has been found to be a potential confound in some instances. Huang, Liu, and Bowling (2015) found in a sample of 345 students, that a prevalence of 10% IER, consistent with both Schmitt & Stults (1985) and Woods (2006), inflated associations between variables in an web-based personality survey and increased Type I error when the response averages of attentive responders were away from modal response options. In an earlier study, Credé (2010) had similar results in a sample of 959 nonacademic employees when measuring depression and neuroticism. This provides the notion that IER can be a potential confound. With the prevalence of IER and the potential impact it can have on data, there have been several indices that have been created to detect IER.

Detection of IER. There a several types of indices that are used to detect IER. The four main approaches of detecting IER are the infrequency, inconsistency, response pattern, and response time approaches (Huang et al., 2012). The infrequency approaches looks at whether or not responders respond to specific items (i.e., bogus items) in a manner that most people would respond in, or when there is clearly a correct response

choice available (Beach, 1989). An example of this is providing response choices of "True" or "False" for the statement "I am the starting quarterback for the Atlanta Falcons". This item would flag respondents for IER if they select "True" because that is unlikely to be true. The inconsistency approach looks at how consistent a responder responds to similar and opposite items. Specifically, there are items on surveys that are either highly positively (psychometric synonym) or negatively correlated (psychometric antonyms) with each other (Goldberg & Kilkowski, 1985) and responses are supposed to be consistent with this. Responses that are inconsistent will be flagged if they hit a certain cutoff point established by the researcher beforehand. The response pattern approach looks at participant responses and determines whether or not their response pattern supports IER. An example of this is if a responder responds to all items with "neutral", this could be a possible indicator of IER because it is highly unlikely that the responder was truly neutral for every response. The response time approach looks at how long responders take to complete a survey and uses the notion that time is negatively correlated with IER. An example of this is if a research participant completes a survey in 10 minutes and the average time to take the survey is an hour than this would be highly indicative of IER. The four different indices have been shown to have good efficacy in detecting IER but have also been found to detect different responders (Meade & Craig, 2012). For instance, the response pattern approach can detect someone who is selecting a response choice frequently but has less efficacy in detecting people who engages in random responding. Essentially this may means that not all IERs have similar response patterns and the efficacy of a particular index is dependent on the type of IER response

pattern. One problem with IER detection indices is that they dichotomize responders as either IER or non-IER.

One flaw in the detection of IER using many of the previously stated indices is that they dichotomize IER. Most research done on the prevalence of IER has shown that, in general, a relatively large portion of responders occasionally engage in IER. Meade and Craig (2012) found that 15% of respondents are flagged by more than one index which indicates that responders occasionally engage in IER. This means that responders can be attentive on one portion of the survey and be inattentive on the next portion. Additionally, Meade and Craig (2012) found that less than 5% of individuals were flagged as engaging in IER at the beginning of surveys but 25% were detected towards the end of the survey. This gives support to the idea that IER is more of a continuous variable, which would make dichotomizing IER inappropriate.

The differences in the reported prevalence, impact, and the efficacy of IER detection indices in detecting IER indicates that there are potentially different types of inattentive responding that may have different prevalence rates and effects on data, which is an area that has yet to be thoroughly explored.

A Distinct Type of IER: IER Due to Survey Fatigue. As previously stated, Meade and Craig (2012) found that IER is more prevalent towards the end of surveys than at the beginning, which is consistent with earlier research (Baer et al., 1997; Berry et al., 1992). One potential explanation for this is survey fatigue. It has been proposed that responders have limited cognitive resources and seek to reduce cognitive effort by engaging in satisficing, a possible manifestation of IER (Krosnick, 1991). The idea behind satisficing is that responders are more likely to retrieve incomplete or biased

information from memory as their cognitive resources are depleted. This researcher also proposed that motivation may play a part in combating satisficing but that over time situational factors (e.g., fatigue) will eventually lead to satisficing. This may be magnified with a long survey.

The length of a questionnaire may cause cognitive fatigue and result in lower response quality. Herzog and Bachman (1981) found that as surveys went on, responders were more likely to respond in a straight line response pattern and that responders were more likely to respond toward modal response options at the end of surveys than at the beginning of a survey. This supports Kraut, Wolfson, and Rothenberg (1975) who found that items placed towards the end of a survey were less likely to endorse extreme responses and more likely to omit responses. These results indicate item positioning could potentially play a factor in response validity, specifically in longer surveys. In support of this notion, Galesic and Bsonjak (2009) had several significant results in their study in which they had participants take a survey that took approximately 30-45 minutes to complete (with "blocks" of items that were randomized to reduce the possibility of confounds). Results indicated that the further away item blocks were from the beginning of the survey the shorter the response times and the lower the response variation to items within those blocks. These studies indicate that items asked later in long surveys bear the risk of producing lower quality data. As support for the theory that IER can be caused by survey fatigue, this study attempted to replicate some of these findings:

Hypothesis 1: Responses to measures placed towards the end of a long survey will have less variance than with responses to the same measure placed towards the beginning of a long survey.

Hypothesis 2a: Responses to measures placed at towards the end of a long survey will be less likely to exhibit extreme responses (e.g., select 1 or 5 on a 5-point Likert Scale) than responses to measures placed towards the beginning of another survey.

Hypothesis 2b: Responses to measures placed towards the end of a long survey will be more likely to endorse central responses choices (e.g., select 3 on a 5-point Likert scale) than responses to the same measure placed closer to towards the beginning of a long survey.

If these results hold true, then there will be further evidence for distinguishing IER due to survey fatigue. Another way to investigate IER due to survey fatigue is observing possible alterations in factor structures from cognitively fatigued respondents.

Merritt (2012) had cognitively fatigued participants take Allen and Meyer's (1990) unidimensional Affective Commitment Scale. The researcher used Merritt and Ilgen's (2008) 30-day trial version of the X-ray Screening Task, which according to Smit et al. (2003) and Warm et al. (2010) are cognitively demanding and have good efficacy in inducing cognitive fatigue (as cited in Merritt, 2012, p.423), to induce cognitive fatigue in participants. Using confirmatory factor analysis, Merritt found in the non-fatigued condition that the unidimensional model or one factor solution worked well. In the fatigue condition, the unidimensional model fit poorly, and that there was a two-factor solution. Furthermore, the model data fit indicated worse fit in the fatigue conditions for the unidimensional model even when negatively worded items were changed to be regularly worded, indicating that fatigue has an effect on factor structures even without reverse worded items. In the same article, Merritt replicated the previous study but in the

replication had an "emphasized" condition in which the researcher modified items to emphasize the contents of negatively worded items (e.g. bolded and underlined the word "Not") and compared these results to an "unemphasized" condition in which items were not modified. Results were similar to the original study and there were no significant differences between the emphasized and unemphasized condition, indicating cognitive fatigued participants may lack the cognitive resources necessary to process negatively worded items even when directed and notified of their contents, which supports Kronick's satisficing theory (1991). Taken as a whole, this study shows that it is possible for cognitive fatigue to induce IER. One limitation of this study is that it only explains the impacts of already cognitive fatigued participants. However, combined with evidence from other studies that found the rate of IER increases towards the end of long surveys (Baer, Ballenger, Berry, & Wetter, 1997; Berry et al., 1992; Meade & Craig, 2011; Meade & Craig, 2012), indicates that long surveys may cause cognitive fatigue and that cognitive fatigue may induce IER. This study investigated whether or not long surveys can induce enough cognitive fatigue to produce similar results.

Hypothesis 3: Measures with negatively worded items placed at the end of a long survey are more likely to have a distinct negatively worded item factors than the same measure placed towards the beginning of a survey.

Hypothesis 4: Items on measures placed at the end of a long survey will be less likely to load on to a latent factor than items on measures placed towards to beginning of a long survey.

Personality Assessments. IER can affect all types of surveys, but one type of commonly used survey at risk of IERs are personality assessments. The early development of IER detection indices were made to detect careless responding in personality tests, in fact, much of the early research on IER was focused on creating detection indices for the commonly used MMPI (Archer & Elkins, 1999; Baer et al., 1997; Buechley & Ball, 1952; Greene, 1978). As mentioned earlier, researchers have been concerned about bad responding in general, but, more attention has been focused on content responsive faking rather than IER. Nichols, Greene, and Schmolck (1989), were some of the first researchers to distinguish IER from content responsive faking and this allowed researchers to separately study these two types of bad responding, which is important because different types of contexts in which surveys are taking may be more at risk of one type of bad responding than the other. For instance, a survey responder may be more likely to fake, rather than engage in IER, on a personality assessment if it is used for as part of the selection process for a job, whereas responders may be more likely to engage in IER, rather than faking, on a personality assessment if it only used for research purposes. With that being said, personality assessments are often used for research purposes and may be particularly vulnerable to IER because they can often be lengthy on their own but are also often included with other items or measures of interest. This is problematic because the as the number of items add up the more impact IER can have on data quality (Credé, 2010). One commonly used personality assessment used in research is the HEXACO-60 (Ashton & Lee, 2009). This manuscript looked at the possible effects of IER on the HEXACO-60.

II. METHODS

To test these hypotheses, this study used archival data consisting of a job analysis survey data from 732 patrol officers from a southeastern highway patrol unit and data from 1008 participants recruited from the Middle Tennessee State University research pool and Amazon's Mechanical Turk program that completed an adaptive performance measure gathered for the purposes of two graduate students' master's theses requirement (of the 1008, 658 of the participants were from a pilot study and 350 were from the actual study). The criteria that needed to be met for the purposes of this study is that two separate studies needed to contain the exact personality measure (i.e., HEXACO-60). These two separate surveys were each split into four sections based on the number of items in the survey (e.g. A 100-item survey would be cut into four separate sections of 25 resulting in a first section containing the first 25 items, a second section containing the second 25 items, etc.). One of the measures needed to have the personality measure at the beginning (i.e. the personality measure needs to be contained within the first section of the survey) and the other survey needed to have the same personality measure towards the end or at the end of the survey (e.g., contained in section 3 or 4). In addition, the survey with the personality measure towards the end of the survey needed to exceed a total item count of 125 and take at least 1-hour to complete. The other items and measures in the respective survey do not have to match one another because the focus of this study was only concerned with how a measure's position may affect response quality to that measure.

Measures

Personality Measure. Ashton and Lee's (2009) HEXACO-60 (refer to Appendix A) will be the personality measure analyzed in this study. This 60-item 5-point scale (1= strongly disagree, 5 = strongly agree) personality instrument is the short form of the 100-item HEXACO-PI-R and measures 6-personality dimensions (i.e., Honest-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience). This personality measure was used as a criteria in selecting archival datasets because of its well documented factor structures, reliability, and internal and external validity (Ashton & Lee, 2009), as well as the availability of datasets that included the HEXACO-60 available to the researcher. In addition, the HEXACO-60 contains reverse scored or negatively worded items, which is necessary to investigate Hypothesis 3. Because of the well documentation of the validation process of the HEXACO-60, responses from the HEXACO-60 in the archival data were analyzed to determine if the actual results deviate from expected results.

The HEXACO-60 was validated using a sample of 936 college students and another sample that was composed of 734 residents from Eugene-Springfield, Oregon (Ashton & Lee, 2009). For the purposes of this study, results from the factor analysis used for validation of the HEXACO-60 were used as a reference and comparison to the results of the factor analysis of this study.

Factor analysis indicated that 37% and 29% of the item variance was accounted by the 6-common factors, respectively of the two samples (Ashton & Lee, 2009). Additional factor analysis of the 24 facets resulted, again, in a 6-factor structure. Each individual item also had primary loadings on the respective factor that was defined by the

other items in its scale. These results provide evidence that the HEXACO-60 measures what it says it measures at the item level. The factor analysis results were used as a comparison for the factor analysis of the HEXACO-60 results in the following two surveys.

Long Survey: For the long survey, data from a job analysis survey from a southeastern state highway patrol unit was analyzed. This survey was used because it contains the HEXACO-60 in section 3 of the survey (right after item number 1533), the total item count is 1754, and took an average of 3 hours to complete. The sample consisted of approximately 732 patrol officers, however only about 526 officers completed the whole survey, and the survey was administered via the online survey platform Qualtrics. The average age of officers was 43 with a standard deviation of 8.66. Of the officers, about 89% were White, 8% were Black, 1% were Hispanic, 1% were Native American and another 1% were Bi-racial/Mixed. In addition, 96% of the officers where male and 4% were female. This survey was intended for job analysis purposes.

Short Survey: For the short survey, data from a measure of individual adaptability was analyzed. This survey (Calarco, 2016; Marlow, 2016) was used because it contains the HEXACO-60 at the beginning of the survey (starts at item 1), and was in section 1 of the survey. There were a total 309 items, and the survey took an average of 30 minutes to complete. Of the 1008 participants recruited from Amazon's Mechanical Turk program and from Middle Tennessee State University's research pool, only 830 completed the survey. The survey was also administered via Qualtrics, an online survey platform, in which the participants who were recruited from Amazon's Mechanical Turk program were paid 20 cents each for completing the survey and research pool participants

received credit in their respective courses. The average age of participants was 37 with a standard deviation of 13.15. Of the participants, 76% were White, 8% were Black, 6% were Hispanic, 4% were Native America, 1% were Asian, and the rest were Mixed. In addition, 31% were male and 69% were females.

Analyses

Item Response Variance. Item response variance was looked at in the HEXACO-60 in the long and short survey. Within-person standard deviations were computed across the HEXACO-60 items and were used to compare between the HEXACO-60 placed at the end of the long survey and the HEXACO-60 placed at the beginning of the short survey. A 0.50 lower *SD* difference in the long survey, compared to the short survey, would indicate there is less response variance in the HEXACO-60 when placed towards to end of a long survey than at the beginning of a survey. A 0.50 *SD* cut off was used because similar cut offs for *SD* difference have been used in past research (Mumford, Schlesinger & Glass, 1982).

Response Count. The frequency of responses to central response categories was calculated to determine if the HEXACO-60 results in the long survey had more responses endorsing central responses (e.g., neutral) than in the short survey.. In addition, the number of extreme responses were compared to determine if the HEXACO-60 measure in the short survey had more responses endorsing extreme responses (e.g. strongly agree) than in the long survey.

Factor Structure. Factor analysis was used to determine whether or not the long survey's HEXACO-60 results had a "negatively worded" factor that appears outside of the 6-personality factors that were expected. In addition, goodness-of-fit indices such as

the Comparative Fit Index (CFI), Tucker-Lewis Incremental Fit Index (TLI), Standardized Root Mean-Square Residual (SRMR) and Root Mean Square Error of Approximation (RMSEA) statistics were used to evaluate the long and short survey's HEXACO-60 model data fit to determine if the long or short survey had more than a 6-factor solution for the HEXACO-60, with higher CFI, TLI, and lower SRMR, and RMSEA indicating better fits. Factor loadings were also analyzed to determine if the HEXACO-60 placed towards the end of the long survey have items that are less likely to load onto their respective factors than the HEXACO-60 placed in the beginning of the short survey.

Procedure. The data from the two separate surveys were cleaned and analyzed using the R statistical environment. Within-person standard deviations and frequency of response categories selected for the HEXACO-60 were calculated and analyzed in the long and short survey and were compared to one another. Additionally confirmatory factor analysis (CFA) was conducted and CFI, TLI, SRMR, and RMSEA statistics were estimated via the Lavaan package (Rosseel, 2012) in the R statistical environment. The results of the CFA of the long and short survey were compared to one another and in addition were compared to the factor structure reported in the validation process of the HEXACO-60.

III. RESULTS

Survey Length and Response Patterns

It was hypothesized that measures placed at the end of a long survey will have different response patterns compared to the same measure placed at the beginning of another survey because of IER. Specifically, it was predicted that the responses to the HEXACO-60 placed at the end of the long survey will have at least a 0.50 *SD* less in response variance for each item compared to the HEXACO-60 placed at the beginning of the short survey (Hypothesis 1), that the HEXACO-60 placed the end of the long survey will have less responses endorsing extreme response categories (i.e., Strongly Disagree or Strongly Agree) compared to the HEXACO-60 placed at the beginning of the short survey (Hypothesis 2a), and that the HEXACO-60 placed at the end of the long survey will have more responses that endorse central response categories (i.e., Neither Agree nor Disagree) than the HEXACO0-60 placed at the beginning of the short survey (Hypothesis 2b).

Response Variance. Hypothesis 1 was tested by computing *SD*s for each item of the HEXACO-60 in both the long and short survey and then subtracting the resulting pair of *SD*s for each item. The results indicated that Hypothesis 1 was not supported in that only one item had at least a 0.50 difference *SD* between the two surveys. Interestingly enough, however, is the fact that even though the differences between the *SD*'s did not reach the predicted SD difference cut off of 0.50, that 59 of the 60 items on the HEXACO-60 when placed at the end of the long survey had less variance than the items on the HEXACO-60 placed at the beginning of the short survey. No additional analyses were conducted (e.g., Chi-square significance test), however, because the small decrease

in *SD* between the long and short survey can be attributed to the fact that the long survey composed of patrol officers in the same southeastern highway patrol unit, while the short survey was composed of participants recruited from Amazon's Mechanical Turk program and Middle Tennessee State University's research pool. Naturally, the participants in the long survey are a more homogenized group compared to the short survey's participants and therefore responses are more likely to be similar to each item - leading to lower response variation in the long survey compared to the short survey.

Central and Extreme Response Endorsement. Responses to the long survey, in regards to endorsing central and extreme response categories, were not different compared to the short survey. Hypothesis 2a and 2b were tested by calculating the frequencies of central and extreme response categories. The results of these calculations were percentages based on the frequency of endorsements for each response category. The percentages for each item and its respective response category percentages were then compared to each other to determine whether or not they differed in in the long and short survey. The results of Hypothesis 2a and 2b were not supported. Of the 60 questions, 35 of the questions had a lower percentage of responders who endorsed "Strongly Agree" and 53 of the questions had a lower percentage of responders who responded "Strongly Disagree" in the long survey compared to short survey, which does not support Hypothesis 2a. Furthermore, of the 60 questions, there were only 34 questions that had a higher percentage of respondents who responded "Neither Agree or Disagree" which does not support Hypothesis 2b. No further analyses (e.g., Chi-square significance test) were conducted because natural difference, as earlier stated, between the long survey

participants and the short survey would most likely explain any of the significant differences in response patterns for the HEXACO-60 in the long and short survey.

Survey Length and Factor Structure

It was hypothesized that placing a measure at the end of a long survey could possible lead to alterations in that measure's factor structure due to IER due to fatigue. Specifically, it was hypothesized that the HEXACO-60 placed at the end of the long survey would have an additional negatively worded item factor (i.e., a seven factor solution) and that the HEXACO-60 placed at the beginning of the short survey would only have a six factor solution (Hypothesis 3). It was also predicted that items on the HEXACO-60 placed at the end of the short survey would be less likely to load onto their respective factors than the same items placed at the beginning of the long survey (Hypothesis 4).

Number of Factors. To test Hypothesis 3 and 4, four confirmatory factor analyses were conducted. Two a priori models were specified in these analyses for both survey datasets. The first a priori model specified was the normal six factor model that Ashton and Lee proposed (2009) and the second, but similar, a priori model specified was a seven factor model. The seven factor model was distinct in that all reverse coded items were loaded onto the additional reverse coded/negatively worded factor. Once the four separate analyses were ran, fit indices and factor loadings were analyzed in order to determine if there was support for Hypothesis 3 and 4.

Refer to Table 1 below for fit indices. Best practices were used in assessing goodness of fit in that an array of fit indices were used rather than just one (Thompson, 2004). It is important to note that interpretation of the fit indices were more emphasized

on the absolute indices of model fit (i.e., RMSEA and SRMR) rather than the relative indices of model fit (i.e., CFI and TLI), which is recommended by Hu and Bentler (1999). Results from the factor analyses fully supported Hypothesis 3. The model data fit for the six factor model was better in the HEXACO-60 placed at the beginning of the short survey, CFI = 0.644, TLI = 0.629, RMSEA = .060, SRMR = .073, than it was in the HEXACO-60 placed at end of the long survey, CFI = 0.579, TLI = 0.561, RMSEA = .064, SRMR = .088, indicating that the six factor model fit worse when the HEXACO-60 was placed at the end of a long survey. Furthermore, the seven factor model fit slightly better than the six factor model when looking at the HEXACO-60 placed at the end of the long survey, CFI = 0.593, TLI = 0.573, RMSEA = .064, SRMR = .085, but, the seven factor model did not fit in the HEXACO-60 placed at the beginning of the short survey, CFI = 0.435, TLI = 0.408, RMSEA = .077, SRMR = .101. This means that Hypothesis 3 was fully supported in that there was a seven factor solution for the HEXACO-60 placed at the end of the long survey but there was not for the HEXACO-60 placed at the beginning of short survey. These results are similar to the findings found in Merritt (2012) and Woods (2006).

Table 1

Model Data Fit for the HEXACO-60 Six and Seven Factor Model to the Long and Short Survey

Model - Data	Fit index				
	CFI	TLI	RMSEA	90% CI RMSEA	SRMR
6 Factor – LS	0.579	0.561	.064	(.063, .066)	.088
7 Factor – LS	0.593	0.573	.064	(.062, .066)	.085
6 Factor – SS	0.644	0.629	.060	(.059, .062)	.073
7 Factor – SS	0.435	0.408	.077	(.075, .078)	.101

Note: LS: Long Survey, SS: Short Survey, CFI: Comparative fit index, TLI: Tucker-Lewes incremental fit index, RMSEA: root mean square error of approximation, SRMR: standardized root mean-square residual

Factor Loadings. Factor loadings were also analyzed to determine if there was evidence for the support of Hypothesis 4. Two things were looked at when analyzing the factor structures: the factor loadings, and the significance of each loading. An alpha of .05 was used for all analyses. Based on this information, Hypothesis 4 was fully supported. Items were more likely to load on their respective factor on the HEXACO-60 when placed at the beginning of the short survey than when the HEXACO-60 was placed at the end of the long survey. This is evident by the fact that, in the six factor model, item "HEX_51" did not load onto Agreeableness, $p = .74$, and item "HEX_50" did not significantly load onto Conscientiousness, $p = .23$, in the long survey but did in the seven factor model. In contrast, all items significantly loaded on their respective factors in the six factor model in the HEXACO-60 placed at the beginning of the short survey. One important to thing to note is that there was a large sample size used for both surveys. This is significant to note in that it is possible that more items may not have not significantly loaded to their respective factors in the longer survey if a smaller size would have been

used. In addition, the item loadings were almost all of a larger magnitude in the HEXACO-60 at the beginning of the short survey than when placed at the end of the long survey. Refer to Tables 2 through 7 on the following pages for the factor loadings.

Table 2

Comparison of Factor Loadings for Honesty and Humility

Factor	Item	Estimate		
Honesty and Humility		Long Survey – Six Factor Model	Long Survey – Seven Factor Model	Short Survey – Six Factor Model
	HEX_6	-0.182*	0.402*	-0.546*
	HEX_12	0.604*	-	0.918*
	HEX_18	-0.125*	-	-0.440*
	HEX_24	0.409*	-	0.384*
	HEX_30	0.435*	-	0.496*
	HEX_36	-0.320*	0.545*	-0.660*
	HEX_42	0.356*	-0.166*	0.623*
	HEX_48	0.491*	-	0.533*
	HEX_54	-0.122*	0.444*	-0.493*
	HEX_60	0.542*	-	0.813*

*$p < .05$

Table 3

Comparison of Factor Loadings for Emotionality

Factor	Item	Estimate		
Emotionality		Long Survey – Six Factor Model	Long Survey – Seven Factor Model	Short Survey – Six Factor Model
	HEX_5	0.542*	0.533*	0.549*
	HEX_11	0.293*	0.291*	0.607*
	HEX_17	0.510*	0.505*	0.621*
	HEX_23	0.522*	0.531*	0.616*
	HEX_29	0.538*	0.528*	0.609*
	HEX_35	-0.116*	-	-0.685*
	HEX_41	-0.256*	-	-0.642*
	HEX_43	0.153*	0.286*	-0.377*
	HEX_49	-0.353*	-	0.764*
	HEX_55	-0.553*	-	0.647*

*$p < .05$

Table 4

Comparison of Factor Loadings for Extraversion

Factor	Item	Estimate		
Extraversion		Long Survey – Six Factor Model	Long Survey – Seven Factor Model	Short Survey – Six Factor Model
	HEX_4	0.375*	0.420*	0.684*
	HEX_10	-0.389*	-	-0.422*
	HEX_16	0.089*	0.342*	0.562*
	HEX_22	0.362*	0.426*	0.710*
	HEX_28	-0.500*	-	-0.662*
	HEX_34	0.188*	0.416*	0.618*
	HEX_40	0.206*	0.401*	0.543*
	HEX_46	-0.468*	-	-0.695*
	HEX_52	-0.570*	-	-0.766*
	HEX_58	0.082*	0.278*	0.494*

*$p < .05$

Table 5

Comparison of Factor Loadings for Agreeableness

Factor	Item	Estimate		
Agreeableness		Long Survey – Six Factor Model	Long Survey – Seven Factor Model	Short Survey – Six Factor Model
	HEX_3	-0.278*	0.675*	0.719*
	HEX_9	0.587*	-	-0.663*
	HEX_15	0.572*	-	-0.566*
	HEX_21	0.613*	-	-0.790*
	HEX_27	-0.296*	0.576*	0.659*
	HEX_33	-0.130*	0.287*	0.534*
	HEX_39	-0.101*	0.300*	0.349*
	HEX_45	-0.226*	0.291*	0.667*
	HEX_51	-0.014	0.253*	0.499*
	HEX_57	0.426*	-	-0.480*

*$p < .05$

Table 6

Comparison of Factor Loadings for Conscientiousness

Factor	Item	Estimate		
Conscientiousness		Long Survey – Six Factor Model	Long Survey – Seven Factor Model	Short Survey – Six Factor Model
	HEX_2	-0.305*	0.466*	-0.652*
	HEX_8	-0.355*	0.513*	-0.373*
	HEX_14	0.486*	-	0.499*
	HEX_20	0.466*	-	0.552*
	HEX_26	0.544*	-	0.671*
	HEX_32	0.572*	-	0.574*
	HEX_38	-0.255*	0.389*	-0.315*
	HEX_44	0.516*	-	0.648*
	HEX_50	-0.055	0.275*	-0.350*
	HEX_56	0.378*	-	0.521*

*$p < .05$

Table 7

Comparison of Factor Loadings for Openness to Experience

Factor	Item	Estimate		
Openness to Experience		Long Survey – Six Factor Model	Long Survey – Seven Factor Model	Short Survey – Six Factor Model
	HEX_1	-0.775*	-	0.815*
	HEX_7	0.500*	0.573*	-0.527*
	HEX_13	0.700*	0.637*	-0.875*
	HEX_19	-0.162*	-	0.412*
	HEX_25	0.768*	0.680*	-0.632*
	HEX_31	-0.445*	-	0.546*
	HEX_37	0.201*	0.313*	-0.510*
	HEX_43	0.153*	0.286*	-0.377*
	HEX_49	-0.353*	-	0.764*
	HEX_55	-0.553*	-	0.647*

*$p < .05$

Supplemental Findings

Total Time and Quality Assurance Items. Additional analyses were conducted that were outside the scope of the previously stated hypotheses. One relationship that was looked at is the relationship between number of quality assurance items answered correctly and total time spent completing the survey. Previous studies have found that IER is negatively correlated with total survey time, meaning careless responders are more likely to spend less time answering items than attentive responders (Huang et al., 2012). The amount of quality assurance items answered correctly in the HEXACO-60 in the short survey and total time spent on the short survey were analyzed to see if there was a significant relationship between the two and if so, whether or not they conform to previous findings on time and IER.

Similar to Meade and Craig' study (2012), the short survey was conducted online, which meant that participants could start the survey, stop, and resume the survey as they wished. This meant that time recorded was not precise, and resulted in many participants who had a total time from start to finish that was over 24 hours. To get around this, only participants who completed the survey within an hour were included in this analysis. The reason for this is approximately 95% of all participants completed the survey within the hour time range. Analysis indicated that total time spent on short survey and number of quality assurance items were positively correlated with each other, $r = .07$, $p < .05$. This means that the longer participants spent on the survey, the more likely they were to answer the quality assurance items correctly. Although a significant relationship was found, a small relationship between numbers of quality assurance items answered correctly and total time spent on the survey may indicate that survey time may not have

good efficacy in detecting IER. One thing to note however, is that the short survey was used and therefore there were only a few responders who did not answer all quality assurance items correctly which may have led to a weaker relationship found.

Reliability. Internal consistency were also compared via Cronbach's alpha. Internal consistency for the HEXACO-60 placed at the end of the long survey ranged from .64 to .78, for each of the 6 personality scales, while the HEXACO-60 placed at the beginning of the short survey ranged from .77 to .82. This means that the HEXACO-60 placed at the end of the long survey was less reliable than when it was placed at the beginning of the short survey.

IV. DISCUSSION

The purpose of this study was to investigate whether or not there would be more evidence of IER in a measure placed at the end of a long survey compared to the same measure placed at the beginning of another survey. This manuscript focused on the idea that one possible type of distinct IER is IER due to survey fatigue and whether or not long surveys have the effect of inducing such inattentive responding. Using the HEXACO-60, this study looked at whether or not IER was more prevalent when the HEXACO-60 was placed at the end of a long survey than when it was placed at the beginning of another survey. This was investigated by comparing the HEXACO-60's response patterns and factor structure of the two surveys.

Response Patterns

There was little evidence to indicate that response patterns were significantly different between the HEXACO-60 placed at the end and beginning of the two surveys. There were slight differences between the two surveys (e.g., the long survey has slightly less response variance compared to the short survey) but many of the differences could be attributed to the fact that the two samples between the two surveys were substantially different. The long survey had a sample comprised entirely of patrol officers from the same southeastern highway patrol unit, whereas the short survey had a sample comprised of the research pool from Middle Tennessee State University and participants who were recruited via Amazon's Mechanical Turk program. Meaning, participants from the long survey were more homogenized, whereas, participants in the short survey were naturally going to be more different from one another and therefore further analyses were not conducted to investigate whether or not response patterns for the HEXACO-60 differed

between the two surveys because the findings would most likely be at risk of confounds. Apart from the response patterns, HEXACO-60 factor structures were also compared between the two studies.

Factor Structures

The HEXACO-60 factor structures were compared between the two studies. Previous research (Merritt, 2012; Schmitt & Stults, 1985; Woods, 2006) has indicated that a presence of IER and reverse coded/negatively worded items have the potential to alter factor structures by introducing an additional distinct reverse coded/negatively worded factor. This study found credence to this in that there was a seven factor solution for the HEXACO-60 placed at the end of the long survey but that there was not in the HEXACO-60 placed at the beginning of a shorter survey. Furthermore, Meade and Craig (2011) suggested that items placed at the end of long surveys may be less likely load onto their respective factors. This was also supported in this study in that there some items that did not significantly load onto their factors in the HEXACO-60 placed at the end of the long survey but this was not the case in the HEXACO-60 placed at the beginning of another short survey, suggesting that long surveys may induce enough IER to alter what items are measuring. This also means that not only do measures placed at the end of long surveys bear a greater risk in producing poorer quality data but also that they bear the risk of altering what items in the measure are actually measuring. Interestingly enough, it was also found that all items that did not load onto their respective factors in the six factor solution, did so in the seven factor solution for the HEXACO-60 placed at the end of the long survey which provides some evidence that the seven factor solution could possibly be used over the normal six factor solution in the long survey. Overall this provides

further evidence that long surveys may lead to enough survey fatigue to induce IER especially when reversed worded items are present.

Supplemental Findings

There were also additional supplemental findings that were of interest. The HEXACO-60 in the short survey had four additional placed quality assurance items. The number of quality assurance items answered correctly and length of time to complete the survey were analyzed to determine if there was a significant relationship between the two variables. Results indicated that the number of quality assurance items answered correctly and time had a very small but significant positive correlation, meaning, participants who took longer on the survey were more likely to answer the quality assurance items correctly. This provides additional evidence for the use of time in detecting IER. Additionally, internal consistency of the HEXACO-60 was assessed for both surveys. Overall, it was found that the HEXACO-60 was more reliable when placed at the beginning of a survey compared to when it was placed at the end of a long survey. This is significant in that the results indicate that survey length may actually decrease reliability after a certain point which is contrary to what is generally believed. With that being said, based on the results of this study, there are additional directions to which future research can take.

Future Research

This current study used archival data to compare data from an applied setting (the long survey) and data from a research pool (the short survey). Naturally, this leads to comparing two different samples that are different from one another which could essentially confound the results of this study which is discussed below. Future research

should collect data from the same population to ensure that results are not the result of the differences between the two samples. Another thing is that the long survey was conducted online, meaning that participants did not have to complete the survey in one sitting. Future studies should determine whether or not the findings in this manuscript can be replicated via a long survey that requires participants to complete it in one sitting. In addition, future research should also use the same survey with the only difference being the measure of interest (in this case the HEXACO-60) placed at the end of the survey or at the beginning. Other personality measures should also be used other than the HEXACO-60 to determine if similar results can be found in other personality measures. Additional studies need to also be done on determining at which point (e.g., number of items) does reliability actually decrease rather than increase for measures placed later in longer surveys. Similarly, individual long measures should be investigated to determine if number of items in one measure can affect reliability for the different scales within that measure. Lastly, this study did not remove any inattentive responders from the HEXACO-60 placed at the end of the long survey; it would be interesting to determine whether or not removing bad responders would affect the factor structure of the HEXACO-60, in other words, does the seven factor solution remain when bad responders are removed? If the seven factor solution does disappear after removing careless responders then this would indicate that the different detection indices are capable of detecting IER but also that the alterations in factor structures are really due to IER. As just mentioned, due to the possible differences between the two samples examined in this study, there are several limitations to keep in mind.

Study Limitations

There are several limitations with this study. One of the major limitations in this study is that the hypotheses are based on the fact that survey participants may become cognitively fatigued from participating in a long survey which in turn leads to IER that affects the psychometric properties of measures. The problem with this is that this study used a long survey conducted that could be completed in multiple sittings. In other words, participants could complete this long survey in small "chunks". This means that there could very well be alternate reasons, other than survey fatigue, that could explain the results.

A second limitation is that participants in the long and short survey are different from each other. The long survey consisted of members of a southeastern state highway patrol unit and this may confound any findings because responders in the long survey may have extra motivation to complete the survey because it may very well affect some aspects of their job which may affect results. Similarly, the short survey included participants who were compensated, either by money or class credit, and therefore may have extra motivation to complete the survey because they're getting compensated to do so, or, they may want to complete the survey as fast as possible to make as much money as possible and may be engaging in IER early in the survey. In other words, this may confound effects because any significant or nonsignificant results may be due to the fact that the two samples taking the HEXACO-60 are different and not be due to survey fatigue.

Lastly, one additional limitation is the possibility of confounds due to the items before the HEXACO-60 in the long survey. Specifically, the long survey is at risk of

possible confounds in that the items before the HEXACO-60 may alter the way participants respond to the HEXACO-60 that is not accounted for by Survey fatigue. Finally another limitation is that the two surveys differ in survey content and ideally would have the same exact measures with the one difference being the positioning of the HEXACO-60. However, this should be a minor limitation because the HEXACO-60 is at the beginning of the short survey so the differences in measures and items in the long and short survey should not have much effect on the results.

The validity of survey responses is important. Without valid responses, the results and the corresponding inferences based on the results are also not valid. Most research done on response validity focuses on faking but less has been done on IER. One way to control for IER is by reducing potential survey fatigue. The results of this study indicate that surveys may need to be constructed to account for survey fatigue and provides additional evidence that IER can affect survey results. Furthermore, it supports the notion that when constructing surveys to only include the measures and items absolutely necessary instead of including additional measures or items for the purposes of self-interest or for use in another research project. Lastly, contrary to the general rule of thumb, the results of this study indicate that longer surveys do not necessarily increase reliability (DeVellis, 1991), in fact, at some point, long surveys may impact the reliability of measures placed later in the survey.

Conclusion

This study contributes to the IER literature in three ways. The main contribution this study makes is that it provides evidence that long surveys can induce enough IER to alter factor structures, which supports previous research done (Merritt, 2012; Schmitt &

Stults, 1985; Woods, 2006). The second contribution of this study was that IER was evident in a dataset from an applied setting (i.e. data from a job analysis conducted for a southeastern highway patrol unit) which provides further evidence that IER is not just found in simulated and student data but is actually problematic in real organizational settings. Lastly, this study provides evidence that there are distinct types of IER. This study found evidence that one form of IER is IER due to survey fatigue. Future research should expand on these findings and find other types of IER. If there is evidence of other distinct types of IER, then researchers can engage in preventative measures by understanding the distinct types of IER when creating surveys in order to prevent the different types of IER from occurring in the first place. Aside from theoretical implications, this study also has some major practical implications.

One of the main practical implications is that survey makers should be cognizant of bad responders, specifically if they have a long survey. This manuscript provides evidence that survey makers have to be concerned with overall survey length in that it may contribute to bad responses which may affect the results of their studies. In addition, this study also shows that careless responders can be a problem in organizational settings and that survey creators in organizations need to account for this by implementing counter measures such as making surveys shorter, adding quality assurance items to the survey, or detecting and removing bad responders by using various post-hoc detection indices.

REFERENCES

Archer, R. P., & Elkins, D. E. (1999). Identification of random responding on the MMPI-A. *Journal of Personality Assessment*, *73*(3), 407-421.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345.

Baer, R. A., Ballenger, J., Berry, D. T., & Wetter, M. W. (1997). Detection of random responding on the MMPI-A. *Journal of Personality Assessment*, *68*(1), 139-151.

Beach, D. A. (1989). Identifying the random responder. *The Journal of Psychology*, *123*(1), 101-103.

Berry, D. T., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*, *4*(3), 340.

Buechley, R., & Ball, H. (1952). A new test of "validity" for the group MMPI. *Journal of Consulting Psychology*, *16*(4), 299.

Calarco, H. L. (2016). Measuring the relationship between adaptive performance and job satisfaction (Unpublished master's thesis). Middle Tennessee State University, Murfreesboro.

Camus, Kelly. Once careless, always careless? Temporal and situational stability of insufficient effort responding (IER). Master's thesis, Wright State University, 2015.

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement*, *70*(4), 596-612.

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.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (Vol. 26). Thousand Oaks, CA: Sage Publications.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method* (Vol. 19). New York, NY: Wiley.

Fleischer, A., Mead, A. D., & Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology*, *8*(2), 196-202.

Furnham, A., Hyde, G., & Trickey, G. (2015). Personality and value correlates of careless and erratic questionnaire responses. *Personality and Individual Differences*, *80*, 64-67.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349-360.

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology*, *48*(1), 82.

Greene, R. L. (1978). An empirically derived MMPI carelessness scale. *Journal of Clinical Psychology*, *34*(2), 407-410.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*(4), 549-559.

Holden, R. R., Wheeler, S., & Marjanovic, Z. (2012). When does random responding distort self-report personality assessment? An example with the NEO PI-R. *Personality and Individual Differences*, *52*(1), 15-20.

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

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828-845.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

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

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality the case of job satisfaction. *Organizational Research Methods*, *18*(3), 512-541.

Kraut, A. I., Wolfson, A. D., & Rothenberg, A. (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology*, *60*(6), 774.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213-236.

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

Marlow, K. K. (2016). The difference between coping and winning: The relationships of adaptive performance, engagement, and conscientiousness (Unpublished master's thesis). Middle Tennessee State University, Murfreesboro.

Meade, A. W., & Craig, S. B. (2011). *Identifying careless responses in survey data*. Paper presented at the 26th Annual Meeting of the Society for Industrial and Organizational Psychology, Chicago, IL.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*. Advance online publication. doi: 10.1037/a0028085

Meade, A. W., & Pappalardo, G. (2013, April). *Predicting careless responses and attrition in survey data with personality*. Paper presented at the 28th Annual Meeting of the Society for Industrial and Organizational Psychology, Houston, TX.

Merritt, S. M. (2012). The two-factor solution to Allen and Meyer's (1990) affective commitment scale: Effects of negatively worded items. *Journal of Business and Psychology*, *27*(4), 421-436.

Mumford, E., Schlesinger, H. J., & Glass, G. V. (1982). The effect of psychological intervention on recovery from surgery and heart attacks: an analysis of the literature. *American Journal of Public Health*, *72*(2), 141-151.

Nichols, D. S., Greene, R. L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, *45*(2), 239-250.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*(4), 367-373.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186-191.

Ziegler, M., MacCann, C., & Roberts, R. (Eds.). (2011). *New perspectives on faking in personality assessment*. New York, NY: Oxford University Press.

APPENDICES

APPENDIX A: HEXACO-60

HEXACO-PI-R

(SELF REPORT FORM)

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

1 ___ I would be quite bored by a visit to an art gallery.

2 ___ I plan ahead and organize things, to avoid scrambling at the last minute.

3 ___ I rarely hold a grudge, even against people who have badly wronged me.

4 ___ I feel reasonably satisfied with myself overall.

5 ___ I would feel afraid if I had to travel in bad weather conditions.

6 ___ I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.

7 ___ I'm interested in learning about the history and politics of other countries.

8 ___ I often push myself very hard when trying to achieve a goal.

9 ___ People sometimes tell me that I am too critical of others.

10 ___ I rarely express my opinions in group meetings.

11 ___ I sometimes can't help worrying about little things.

12 ___ If I knew that I could never get caught, I would be willing to steal a million dollars.

13 ___ I would enjoy creating a work of art, such as a novel, a song, or a painting.

14 ___ When working on something, I don't pay much attention to small details.

15 ___ People sometimes tell me that I'm too stubborn.

16 ___ I prefer jobs that involve active social interaction to those that involve working alone.

17 ___ When I suffer from a painful experience, I need someone to make me feel comfortable.

18 ___ Having a lot of money is not especially important to me.

19 ___ I think that paying attention to radical ideas is a waste of time.

20 ___ I make decisions based on the feeling of the moment rather than on careful thought.

21 ___ People think of me as someone who has a quick temper.

22 ___ On most days, I feel cheerful and optimistic.

23 ___ I feel like crying when I see other people crying.

24 ___ I think that I am entitled to more respect than the average person is.

25 ___ If I had the opportunity, I would like to attend a classical music concert.

26 ___ When working, I sometimes have difficulties due to being disorganized.

27 ___ My attitude toward people who have treated me badly is “forgive and forget”.

28 ___ I feel that I am an unpopular person.

29 ___ When it comes to physical danger, I am very fearful.

30 ___ If I want something from someone, I will laugh at that person's worst jokes.

31 ___ I’ve never really enjoyed looking through an encyclopedia.

32 ___ I do only the minimum amount of work needed to get by.

33 ___ I tend to be lenient in judging other people.

34 ___ In social situations, I’m usually the one who makes the first move.

35 ___ I worry a lot less than most people do.

36 ___ I would never accept a bribe, even if it were very large.

37 ___ People have often told me that I have a good imagination.

38 ___ I always try to be accurate in my work, even at the expense of time.

39 ___ I am usually quite flexible in my opinions when people disagree with me.

40 ___ The first thing that I always do in a new place is to make friends.

41 ___ I can handle difficult situations without needing emotional support from anyone else.

42 ___ I would get a lot of pleasure from owning expensive luxury goods.

43 ___ I like people who have unconventional views.

44 ___ I make a lot of mistakes because I don't think before I act.

45 ___ Most people tend to get angry more quickly than I do.

46 ___ Most people are more upbeat and dynamic than I generally am.

47 ___ I feel strong emotions when someone close to me is going away for a long time.

48 ___ I want people to know that I am an important person of high status.

49 ___ I don't think of myself as the artistic or creative type.

50 ___ People often call me a perfectionist.

51 ___ Even when people make a lot of mistakes, I rarely say anything negative.

52 ___ I sometimes feel that I am a worthless person.

53 ___ Even in an emergency I wouldn't feel like panicking.

54 ___ I wouldn't pretend to like someone just to get that person to do favors for me.

55 ___ I find it boring to discuss philosophy.

56 ___ I prefer to do whatever comes to mind, rather than stick to a plan.

57 ___ When people tell me that I'm wrong, my first reaction is to argue with them.

58 ___ When I'm in a group of people, I'm often the one who speaks on behalf of the group.

59 ___ I remain unemotional even in situations where most people get very sentimental.

60 ___ I'd be tempted to use counterfeit money, if I were sure I could get away with it.

APPENDIX B: Date Release Form

DATA RELEASE AGREEMENT

Colonel Tracy Trott agrees to make available to **Hung Loan Nguyen** and **Megan Wertheimer** the following:

Data requested: **THP Job Analysis Survey data**

Purpose for which the data are to be used:

To be analyzed and interpreted in order to complete a thesis for the purposes of the partial fulfillment of the requirements for the degree of Master of Art in Industrial Organizational Psychology.

Conditions:

Hung Loan Nguyen and **Megan Wertheimer** agrees that the data will not be released in whole or in part to any individual or organization, other than the thesis advisor, **Dr. Michael Hein**, without prior written consent from THP. This restriction applies to all reorganizations of the data, in whole or in part, and to integrations of the data with information from other sources.

Hung Loan Nguyen and **Megan Wertheimer** agrees to clearly acknowledge the source of the data supplied by THP whenever such data are used in any report, publication, document or public communication.

The data provided may only be used in reports or presentations directly related to the purpose described above.

THP makes no warranties as to the accuracy of the data nor its suitability for **Hung Loan Nguyen** and **Megan Wertheimer**'s purpose.

THP does not guarantee exclusivity of use of the data.

It is understood that **Hung Loan Nguyen** and **Megan Wertheimer** will destroy all electronic or paper copies of the data (excluding products generated from the data such as reports, maps, documents or public communications) at the termination of this agreement. Data will remain housed on Dr. Hein's computer.

By signing this agreement, and accepting the data, **Hung Loan Nguyen** and **Megan Wertheimer** agrees to be bound by the above conditions.

9/8/16

Hung Loan Nguyen Date

9/8/16

Megan Wertheimer Date

Col TC Trott 9/7/16

Tracy Trott Date

APPENDIX C: IRB Approval Form

IRB
INSTITUTIONAL REVIEW BOARD
Office of Research Compliance,
010A Sam Ingram Building,
2269 Middle Tennessee Blvd
Murfreesboro, TN 37129

MIDDLE TENNESSEE STATE UNIVERSITY

IRBN007 – EXEMPTION DETERMINATION NOTICE

Monday, September 26, 2016

Investigator(s): Hung Loan Nguyen (Student PI), Michael Hein (FA) and Mark Frame
Investigator(s') Email(s): *htn2h@mtmail.mtsu.edu; michael.hein@mtsu.edu; mark.frame@mtsu.edu*
Department: Psychology

Study Title: ***Tired of survey fatigue? Insufficient effort responding due to survey fatigue***
Protocol ID: **17-1028**

Dear Investigator(s),

The above identified research proposal has been reviewed by the MTSU Institutional Review Board (IRB) through the **EXEMPT** review mechanism under 45 CFR 46.101(b)(2) within the research category *(4) Study involving existing data* A summary of the IRB action and other particulars in regard to this protocol application is tabulated as shown below:

IRB Action	EXEMPT from furhter IRB review***	
Date of expiration	NOT APPLICABLE	
Sample Size	1,000 (ONE THOUSAND)	
Participant Pool	**Adult (18 and older)**	
Mandatory Requirements	Not eligible to recruit participants	
Additional Restrictions	NONE	
Comments	NONE	
Amendments	Date	**Post-Approval Amendments** NONE

***This exemption determination only allows above defined protocol from further IRB review such as continuing review. However, the following post-approval requirements still apply:

- Addition/removal of subject population should not be implemented without IRB approval
- Change in investigators must be notified and approved
- Modifications to procedures must be clearly articulated in an addendum request and the proposed changes must not be incorporated without an approval
- Be advised that the proposed change must comply within the requirements for exemption
- Changes to the research location must be approved – appropriate permission letter(s) from external institutions must accompany the addendum request form
- Changes to funding source must be notified via email (irb_submissions@mtsu.edu)
- The exemption does not expire as long as the protocol is in good standing
- Project completion must be reported via email (irb_submissions@mtsu.edu)

- Research-related injuries to the participants and other events must be reported within 48 hours of such events to compliance@mtsu.edu

The current MTSU IRB policies allow the investigators to make the following types of changes to this protocol without the need to report to the Office of Compliance, as long as the proposed changes do not result in the cancellation of the protocols eligibility for exemption:

- Editorial and minor administrative revisions to the consent form or other study documents
- Increasing/decreasing the participant size

The investigator(s) indicated in this notification should read and abide by all applicable post-approval conditions imposed with this approval. Refer to the post-approval guidelines posted in the MTSU IRB's website. Any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918 within 48 hours of the incident.

All of the research-related records, which include signed consent forms, current & past investigator information, training certificates, survey instruments and other documents related to the study, must be retained by the PI or the faculty advisor (if the PI is a student) at the sacure location mentioned in the protocol application. The data storage must be maintained for at least three (3) years after study completion. Subsequently, the researcher may destroy the data in a manner that maintains confidentiality and anonymity. IRB reserves the right to modify, change or cancel the terms of this letter without prior notice. Be advised that IRB also reserves the right to inspect or audit your records if needed.

Sincerely,

Institutional Review Board
Middle Tennessee State University

Quick Links:

Click here for a detailed list of the post-approval responsibilities.
More information on exmpt procedures can be found here.