2014

The Effectiveness of Warnings at Reducing the Prevalence of Insufficient Effort Responding

Caitlin E. Blackmore
Wright State University

THE EFFECTIVENESS OF WARNINGS AT REDUCING THE PREVALENCE OF INSUFFICIENT EFFORT RESPONDING

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

By

CAITLIN ELISE BLACKMORE

B.S., Texas State University, 2010

2014

Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DATE: APRIL 30, 2014

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Caitlin Blackmore ENTITLED The Effectiveness of Warnings at Reducing the Prevalence of Insufficient Effort Responding BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

Nathan Bowling, Ph.D.
Thesis Director

Scott N. J. Watamaniuk, Ph.D.
Graduate Program Director

Debra Steele-Johnson, Ph.D.
Chair, Department of Psychology

Committee on
Final Examination

Nathan Bowling, Ph.D.

David LaHuis, Ph.D.

Gary Burns, Ph.D.

Robert E. W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

ABSTRACT

Blackmore, Caitlin Elise. M.S., Department of Psychology,
Industrial and Organizational Psychology Program, Wright State University, 2014.
The Effectiveness of Warnings at Reducing the Prevalence of Insufficient Effort Responding.

Unmotivated participants who fail to devote sufficient effort to their survey responses can influence the quality of self-report data. The majority of the published literature on this topic has concerned techniques for detecting insufficient effort responding (IER), whereas little attention has been given to developing effective procedures for preventing IER. There are numerous advantages to preventing IER, one of which is that discarding data is unnecessary. The current study examined the effects of a warning manipulation on the prevalence of IER and the quality of the resulting data. Statistically significant differences between conditions on four of the IER detection measures were observed, indicating a lower prevalence of IER in the warning condition compared to the control condition. Results also supported existing research on IER's consequences to internal consistency reliability, and demonstrated that a warning manipulation can impact internal consistency reliability estimates. These findings have important implications relevant to both researchers and practitioners who collect self-report data, as they demonstrate the efficacy of an IER intervention that has numerous advantages over the more commonly used IER detection measures.

TABLE OF CONTENTS

LIST OF TABLES AND APPENDICES

I. INTRODUCTION

Careless responses to items on a questionnaire, also known as Insufficient Effort Responding (IER;Huang, Curran, Keeney, Poposki, & Deshon, 2012), can negatively impact the psychometric properties of measures (e.g., Huang et al., 2012; Meade & Craig, 2012; Woods, 2006) as well as the validity of observed interconstruct correlations (Huang, Bowling, Liu, & Li, In press; Huang, Liu, & Bowling, In review). Researchers have estimated that about 10% of a given sample will respond inattentively (e.g., Huang et al., In Press; Kurtz & Parish, 2001; Meade & Craig, 2012), which appears to be sufficient to generate these negative consequences (e.g., Huang et al., In Press; Schmitt & Stults, 1985; Woods, 2006). However, few researchers have examined IER despite its clear practical importance (Liu, Bowling, Huang, & Kent, 2013). The investigation of this topic has become increasingly important due to the growing popularity of online questionnaires, which are characterized by three of the factors that have been identified as contributing to IER: Anonymity, lack of personalization between the researcher and participant, and environmental distraction (see Meade & Craig, 2012). Fortunately, researchers have recently begun to devote more attention to methods of detecting IER (e.g. Johnson, 2005; Meade & Craig, 2012; Huang et al., 2012); however, little consideration has been given to the prevention of IER. The ability to prevent IER is of great practical importance and has many benefits over IER detection, the most notable of which is that the removal of data is unnecessary. In the current study, I will evaluate the effectiveness with which warnings reduce the prevalence and associated negative consequences of IER.

Types of Responding

Nichols, Greene, and Schmolck (1989) identified two ways in which respondents can provide poor quality survey data: Content responsive faking and content nonresponsivity. Content responsive faking is characterized by participants attending to item content but providing inaccurate responses, which includes both intentional and unintentional faking. This type of responding has been extensively studied (e.g., Paulus, 1984; Viswesvaran & Ones, 1999) and will not be the focus of the current research. Alternatively, Nichols et al. (1989) described content nonresponsivity as occurring when participants respond to survey items without consideration of item content. Although most researchers have referred to this phenomenon as *random* responding in the literature (e.g., Beach, 1989; Berry, Wetter, Baer, Larson, Clark, & Monroe, 1992; Charter, 1994), this label is somewhat misleading. Some participants, for example, employ the strategy of providing the same response option to many successive items, which is clearly not random. In order to provide a more accurate and inclusive description, Huang et al. (2012) proposed the label *Insufficient Effort Responding* (IER), "a response set in which the respondent answers a survey measure with low or little motivation to comply with the survey instructions, correctly interpret item content, and provide accurate responses" (pp. 100).

Consequences of IER

IER can have numerous consequences to the quality of self-report data. Specifically, researchers have noted that IER can influence internal consistency reliability (Huang et al., 2012), predictor-criterion relationships (Huang et al., In press), and the factor structures of measures (e.g., Huang et al., 2012; Meade & Craig, 2012; Woods,

2006). Concern over the quality of self-report data applies to any discipline that utilizes questionnaires to collect data; however, it is especially relevant to the social and behavioral sciences, including Industrial-Organizational Psychology, where surveys are the most common method of data collection (Malhorta, Kim, & Patil, 2006; Spector, 2002).

Internal Consistency Reliability and Interconstruct Correlations. Although it seems intuitive that IER would attenuate internal consistency reliability, the opposite can occur if enough inattentive respondents choose the same response option repeatedly on a scale that is primarily scored in one direction (Huang et al., 2012; Huang et al., In press). There are several factors that contribute to the way in which IER affects observed relationships between substantive variables. Firstly, if IER contamination is present in both the predictor and criterion measures, it is essentially common method variance and will result in an artificially inflated observed correlation between the substantive variables. However, if IER contamination is present in only one of the measures, it produces a suppressor effect and will result in an artificially attenuated observed correlation (Conway & Lance, 2010). Huang et al. (In Press) also found evidence that IER attenuates relationships between scales of the same construct when the items of each scale are worded in opposite directions, and inflates relationships between different constructs when the items of each scale are worded in the same direction. Additionally, Huang et al. (In Review) recently made the observation that inattentive respondents collectively score around the midpoint of the response scale on substantive variables. This has important implications for the effects of IER on observed correlations. The authors refer to the mean score of attentive participants relative to the midpoint as "D,"

and specify that a high mean of attentive participants relative to the midpoint signifies a negative D, and a low mean of attentive participants relative to the midpoint signifies a positive D. Using simulated data, the authors found that when the Ds of two substantive variables were in the same direction, IER created a positive correlation where one should not have existed. Alternatively, when the Ds of two substantive variables were in opposite directions, IER created a negative correlation where one should not have existed. It seems reasonable to predict that a similar phenomenon would occur if the two substantive variables were correlated, but instead of creating a positive or negative correlation where one did not exist, IER would either inflate or attenuate the true correlation. However, an additional factor influencing IER's effects on predictor-criterion relationships is introduced by the existence of a correlation between two substantive variables: the slope of the regression line for the relationship between the substantive variables. If the slope of the regression line is such that the midpoint of the response scales—and thus the mean score of inattentive participants—lies on the regression line, IER will either inflate or have no effect on the observed correlation. Alternatively, if the slope of the regression line is such that the midpoint of the response scales lies some distance from the regression line, IER will attenuate the observed correlation. This should occur regardless of the valence of the Ds, which greatly limits the accuracy with which researchers can predict the effects of IER on predictor-criterion relationships.

In summary, the effects of IER on interconstruct correlations should be unpredictable due to the complex mechanisms that influence these effects. Unpredictable biases are unfortunately especially detrimental to research findings and subsequent

conclusions due to our inability to prevent them or correct for them. As a part of this study, I will attempt to replicate Huang et al.'s (In press) findings regarding the specific effects of IER on internal consistency reliability estimates, and I will document the effects of IER on interconstruct correlations. In addition, I will examine the effects of a warning manipulation on internal consistency reliability estimates and observed interconstruct correlations.

Factor Analytic Results. Researchers have also found that IER can influence the factor structures of measures by producing "method" factors (Huang et al., 2012; Schmitt & Stults, 1985; Woods, 2006). Method factors occur when positively worded items and negatively worded items from one scale load onto two separate factors. These studies demonstrated that a single-factor solution, which would have been the hypothesized solution assuming the measures were adequate, did not fit the data until IER cases were identified and removed. Furthermore, this effect occurred when as little as 10% of the sample provided careless responses.

Context Effects. Another consequence that IER might have is to prevent context effects from occurring. When referring to surveys, researchers generally define context effects as how preceding questions on a survey affect participants' responses to subsequent questions (e.g., Schuman & Presser, 1981; Schuman, 1992; Sudman, Bradburn, & Schwarz, 1996). Based on this definition, it is easy to see how IER could influence the manifestation of context effects. If a participant does not read the survey items, it is unlikely that the participant's responses will be affected by preceding questions. Bowling, Burns, Huang, and Blackmore (In preparation) developed a method of measuring the presence of these context effects, which involves planting items within a

survey. These items are designed to prime the participant to respond to a later, open-ended question with the same ideas that were presented in the embedded items. The subjects of the priming items should not be reflected in inattentive participants' responses to the open-ended questions. The authors do not see this method as an approach to IER detection, but rather as a way to document another consequence of IER and provide additional evidence of construct validity. This study will use Bowling et al.'s (In preparation) method to examine the effects of IER and a warning manipulation on the occurrence of context effects.

Prevalence of IER

Although estimates of the prevalence of IER (e.g., Curran Kotrba, & Denison, 2010; Ehlers, Greene-Shortridge, Weekly, & Zajack, 2009; Johnson, 2005; Kurtz & Parish, 2001; Meade & Craig, 2012) have ranged from 3.5% (Johnson, 2005) to 72% (Baer, Ballenger, Berry, & Wetter, 1997), the incidence in a typical sample appears to be sufficient to produce the negative consequences of IER. Both Schmitt and Stults (1985) and Woods (2006), for example, used simulated data to show that factor structures could be distorted when just 10% of a sample was responding inattentively. Similarly, Huang et al. (2012) found that scale psychometric properties were significantly improved by excluding data provided by only a small number of inattentive respondents.

The disparity in IER prevalence estimates is due to both the method and the criteria used to identify IER. Curran et al. (2010) found that anywhere from 5% to 50% of their sample responded carelessly to survey items depending on the method employed and the cutoff criteria used to define IER. Meade and Craig (2012) made the observation that the methods used to detect IER most effectively identify the few participants who

provide inattentive responses throughout much of the survey, and are less effective at detecting occasional IER. Unfortunately, research has shown that although the majority of participants do engage in IER, most will only provide inattentive responses periodically. For example, 50%-60% of Berry et al.'s (1992) sample, and 73% of Baer et al.'s (1997) sample, admitted to providing a small number of inattentive responses. Meade and Craig (2012) and Huang et al. (In Press), on the other hand, found that only about 10% to 12% of participants provided a large number of inattentive responses.

Practical Importance of IER

The prevalence of IER and the many negative consequences that have been attributed to IER are testaments to the importance of this topic. Although IER is not as likely to be an issue among job applicants due to the high-stakes nature of the situation, this issue still has practical significance to other areas of personnel selection. For example, job incumbents often are used to collect data for job analysis, concurrent validation, and survey feedback interventions. Because job incumbents have less motivation to attend to organizational questionnaires than job applicants, some job incumbents are likely to provide inattentive responses, especially if anonymous surveys are used to collect this data. This could have serious consequences, as the results of these studies are frequently used to design organizational interventions that have the potential to affect employees' lives.

In spite of the clear importance of this topic, few researchers have made it a priority to investigate IER (Liu et al., 2013). In addition, although many researchers do advise cleaning data before analysis, the recommended and commonly used methods focus on missing data, outliers, and skewed distributions, and are insufficient for the

purpose of identifying IER. Unfortunately, no standardized IER detection procedure is currently in existence, so researchers have been forced to resort to more ad hoc methods, such as visual inspection of unusual response patterns.

Factors that Contribute to IER

One reason for the lack of research attention given to this topic may be that the majority of researchers assume that IER is relatively uncommon (Johnson, 2005; Liu et al., In Press). However, research has shown that data quality is affected even when very few inattentive responses occur in a sample (e.g., Huang et al., 2012; Schmitt & Stults, 1985; Woods, 2006). Additionally, there is reason to believe that IER is increasing due to the growing use of Internet surveys (Simsek & Veiga, 2001) and undergraduate samples (Peterson, 2001). Many universities in the United States require undergraduate students in Introductory Psychology courses to participate in research for course credit. Although this is convenient for researchers, the fact that participation is a course requirement means that students may not be sufficiently motivated to provide quality data. Online surveys and anonymity can exacerbate this effect. Beach (1989) found that participants who completed an online survey were significantly more likely to engage in IER than participants who completed a paper-and-pencil survey. Research has also shown that online anonymity tends to produce a lack of accountability (Douglas & McGarty, 2001; Lee, 2006), which could encourage some participants to respond carelessly to surveys.

Meade and Craig (2012) identified three additional factors that affect the likelihood of participants providing inattentive responses. The first of these factors is survey length. Very long surveys can deplete an individual's resources, increasing the

probability that even the most motivated participant will feel fatigued and provide inattentive responses. Supporting this idea, research has demonstrated that IER is positively related to survey length (Baer et al., 1997; Berry et al., 1992). The second factor that has the potential to evoke inattentive responses is lack of personalization, which refers to the social distance between the researcher and participant. Similar to the effects of anonymity, a lack of personalization may cause a participant to feel less accountable for their actions, and thus less motivated to provide attentive responses (Johnson, 2005). Unfortunately, this phenomenon is an unavoidable consequence of online surveys. The third factor influencing the likelihood of IER, and yet another problem inherent to online surveys, is environmental distraction. Because the researcher rarely has control over the setting in which participants complete online surveys, there is a chance that some participants will choose distracting environments or attempt to multitask while responding to the survey. Research demonstrating the detriments to performance associated with divided attention (Spelke, Hirst, & Neisser, 1976) suggests that these participants would be more likely to engage in IER.

Approaches to Detecting IER

Perhaps driven by the perception that IER is a growing problem, one of the recent objectives of IER researchers has been examining the relative effectiveness of the various IER detection methods that have been developed. There are five common approaches to IER detection: 1) the infrequency approach, 2) the inconsistency approach, 3) the response pattern approach, 4) the response time approach (Huang et al. 2012), and 5) the self-report approach (Meade & Craig, 2012), as well as one newly developed approach: the item content recognition approach (ICRA). The methods subsumed under these

approaches use one of two tactics: they involve either planting items designed to detect IER within the survey or performing statistical analyses after data collection (referred to as "post-hoc" methods by Meade & Craig, 2012). In the following sections, I will discuss the advantages and disadvantages of each of the approaches to IER detection.

The Infrequency Approach. The infrequency approach to detecting IER includes several measures consisting of special items that the scale developers designed to elicit the same response from all attentive participants. Researchers who use these scales embed the special items amongst the substantive items in a questionnaire, and assess the likelihood that a participant engaged in IER by the number of anomalous responses that they provided. The infrequency approach includes instructed response items (e.g., "Please select strongly disagree for this question") and nonsense items, such as Meade and Craig's (2012) "Bogus" Item Scale (e.g., "I have been to every country in the world") and Huang et al.'s (In press) IER Scale (e.g., "I work twenty-eight hours in a typical workday"). Bowling et al. (In preparation) found a correlation of 0.90 between the Bogus Item Scale and the IER Scale, indicating that they are essentially parallel measures. In a recent examination of the efficacy of IER detection methods, Meade and Craig (2012) found that their Bogus Item Scale was one of the most effective methods assessed. Although this finding is encouraging, there are some disadvantages of this method that are worth mentioning. One is that adding these items increases the length of the survey, which can contribute to participant fatigue and evoke the very behavior we are concerned with eliminating. Also, because this method involves adding items to the survey, it cannot be used when involvement in the survey development process is not possible. Another concern specific to nonsense items is that they may influence

participants' responses to the substantive items in the survey. For example, some participants may conclude from the item "I am paid biweekly by leprechauns" (Meade & Craig, 2012), that the researchers or study is not credible, which may decrease motivation to provide valid responses.

The Inconsistency Approach. The inconsistency approach is based on the idea that attentive participants will respond to conceptually similar items consistently across a survey. The most common IER detection indices that take this approach are the psychometric synonym and antonym indices (Goldberg, 1985) and the individual reliability index (Jackson, 1976). Psychometric synonyms are items that are conceptually equivalent (e.g., "Carry out my plans" and "Finish what I start"), and should therefore be positively correlated, and psychometric antonyms are items that are conceptually antagonistic (e.g., "Dislike myself" and "Am very pleased with myself"), and should therefore be negatively correlated. These indices are traditionally computed by examining the inter-item correlations between all of the items in a survey. Item pairs with correlations of .60 or greater serve as the psychometric synonym index, and item pairs with correlations of -.60 or less serve as the psychometric antonym index (Meade & Craig, 2012). Within-person correlations between each set of items are then computed, and these correlations constitute the scores for each index, with possible scores ranging from -1 to 1. The more positive a participant's psychometric synonym score, the less likely it is that they engaged in IER. Similarly, the more negative a participant's psychometric antonym score, the less likely it is that they engaged in IER. Although Huang et al. (2012) and Meade and Craig (2012) found support for the usefulness of this method, Meade and Craig (2012) also noted that its efficacy was heavily contingent on

the number of item pairs comprising the index. Goldberg (1985) suggested using thirty item pairs, which limits the use of this method to very long surveys that yield at least thirty highly correlated item pairs per index.

Individual reliability is computed by splitting each unidimensional scale in a survey into an even subscale and an odd subscale, using the item numbers based on order of presentation. A score is then computed for each subscale by averaging the responses, and the correlation between the even and odd subscales—corrected with the Spearman-Brown split half formula—is computed for each participant. The even and odd subscale scores should be highly correlated among attentive participants, allowing researchers to identify participants who may have been engaging in IER. Both Huang et al. (2012) and Meade and Craig (2012) found support for the efficacy of the individual reliability index; however, similar to the psychometric synonym and antonym indices, a large number of scales (e.g., 30) are necessary to produce a stable correlation coefficient. Another disadvantage of this method is that it is insensitive to occasional inattentive responses due to its reliance on subscale scores rather than individual items (Meade & Craig, 2012).

There are two additional disadvantages that are common to all of the methods comprising the inconsistency approach. The first of these is that response consistency is an individual difference. Goldberg (1985) explained that some participants use both counts and intensities of past behaviors to determine how to respond to self-report items. Furthermore, participants may consider their behavior in different contexts when responding to different items, or they may average their behavior over the different contexts. This can result in inconsistent responses that do not reflect IER, which suggests that using the inconsistency indices to detect IER may be problematic. The second

problem with the inconsistency approach methods is that they are insensitive to nonrandom IER response patterns, which are produced when participants select the same response option for multiple consecutive items (Meade & Craig, 2012). This response strategy yields highly consistent responses with little variance. Because the methods constituting the inconsistency approach are a measure of inconsistent (or random) response patterns, and thus require variance in responses, they are unable to effectively detect nonrandom IER response patterns.

The Response Pattern Approach. The response pattern approach offers a solution to the problem of identifying nonrandom IER response patterns. This approach involves identifying the longest string of identical responses, or the greatest number of times a participant provided the same response option to consecutive items across the survey (Costa & McCrae, 2008). Meade and Craig (2012) referred to this as the "maximum long string index," and also computed an "average long string index" by taking the average of the longest string of identical responses for each survey page. Because this approach is specifically designed to identify responses that are too consistent, it is unable to effectively detect inconsistent (i.e., random) IER response patterns. Unfortunately, the evidence thus far suggests that most participants who engage in IER provide random, as opposed to nonrandom, response patterns (Meade & Craig, 2012), which significantly limits the usefulness of this approach.

The Response Time Approach. The response time approach typically utilizes timestamps generated by computer-based surveys to record the amount of time that a participant spent responding to the survey items. Two common operational representations of response time are the total time spent responding to the survey and the

average time spent on each page of the survey. Huang et al. (2012) provided evidence in favor of the average page time measure of IER, reporting that it was one of the most effective methods examined. However, the fact that this approach generally assumes that inattentive respondents can have either very short or very long response times (Meade & Craig, 2012) suggests that the interpretation of response time may be problematic. It is probably safe to use this method to eliminate only the participants whose response times are so short that it would be virtually impossible to attentively respond to all of the items. Using the distribution of response times to determine an appropriate cutoff, however, is problematic. There are many explanations beyond IER for variability in response time. In fact, psychologists have used response time to study various cognitive processes (e.g., Fazio & Williams, 1986; Smith & Miller, 1983). Therefore, there is no way to be certain that relatively short or long response times are attributable to IER. Additionally, even attentive respondents may have very long response times if they complete the survey in multiple sessions or are interrupted while completing the survey.

Self-Reported IER. A relatively new strategy for detecting IER is to ask participants directly how much effort they allocated to the survey. Meade and Craig (2012) developed several self-reported IER scales, including the Diligence scale. This measure consists of eight items that ask participants the extent to which they provided thoughtful responses (e.g., "I carefully read every survey item."). Although the convenience and simplicity of this strategy is appealing, there are obvious problems with attempting to obtain valid self-report data from participants who are not reading the survey items or responding attentively. Preliminary evidence would seem to support this concern; researchers of the two studies to date that have employed self-reported IER

scales (Huang et al., 2012; Meade & Craig, 2012) concluded that the self-report items alone were not sufficient to detect IER.

Item Content Recognition Approach (ICRA). Bowling et al. (In preparation) developed and validated a new approach to IER detection called the item content recognition approach (ICRA), which consists of memorable "target" items that are evenly distributed throughout a survey (e.g. "I have had a recurring dream in which all my teeth have fallen out."), and "quiz" items on the final page of the survey that test participants over the content of the target items. The authors have found evidence in support of the validity of the ICRA. In addition, they believe that the method should have the ability to detect both random and nonrandom IER response patterns because the quiz items utilize more than one correct response option.

Item Response Theory. Although it is beyond the scope of the current study, it is also worth mentioning that item response theory-based measures have also been applied to IER detection. For example, Lumsden (1977) explained how a person response function (PRF), or the relationship between item difficulty and the proportion of correctly answered dichotomous items for each person, could be used to determine the extent to which participants' responses are consistent with the IRT model.

Approaches to Preventing IER

Another recent development in this stream of research is the idea that IER can be prevented. So far, researchers have only investigated two IER prevention methods: warnings and identity disclosure. In one study, researchers warned a sample of undergraduate students that "sophisticated statistical control methods" (pp. 103) would be able to detect inattentive responses, and that participants would lose credits if they were

identified by these methods (Huang et al., 2012). The authors' results indicated that warnings significantly reduced IER as identified by psychometric antonyms, individual reliability, and the long string index.

In another study, researchers used the finding that online anonymity reduces accountability (Douglas & McGarty, 2001; Lee, 2006) as the rationale to incorporate a condition in which participants were required to disclose their identities (Meade & Craig, 2012). The authors also included a warning condition, in which they told participants, "Remember that your honesty and thoughtful responses are subject to the university's academic integrity policy" (pp. 441) and required participants to subsequently sign a statement that read, "I verify that I have carefully and honestly answered all questions on this page in accordance with the university's honor policy" (pp. 441). Contrary to Huang et al.'s (2012) results, Meade and Craig (2012) found no significant differences between the warning condition and the control condition for any of the IER detection methods utilized. However, Meade and Craig (2012) did find that identity disclosure significantly reduced the number of bogus items endorsed and resulted in significantly higher scores on the self-reported attention scale. These two studies are unique in their investigation of IER prevention, and unfortunately, their contradictory findings shed little light on whether or not prevention is a feasible way to contend with IER. Thus, the primary purpose of the current study will be to further investigate the effectiveness of warnings as an IER prevention method.

Unanswered Questions and Study Objectives

Effectiveness of IER Prevention. One explanation for the disparities in Meade and Craig's (2012) and Huang et al.'s (2012) results can be found in the literature

examining faking on personality assessments. Dwight and Donovan (2003) reviewed past studies on warnings to not fake, and noted that three types of warnings can be utilized: warnings that communicate the ability of the test administrators to identify faking, warnings that communicate the consequences of being identified as faking, and warnings that communicate both the ability to identify faking and the consequences associated with faking. The authors performed a basic meta-analysis of studies that had examined the effects of warnings on faking, and found that the type of warning administered moderated the relationship between warnings and faking. Specifically, estimates of effect sizes were larger for studies that utilized warnings communicating a consequence and studies that utilized warnings communicating both a consequence and the ability of the test administrators to identify faking. The authors then performed an experiment comparing the relative efficacy of each type of warning in reducing faking on a personality measure. Supporting the authors' meta-analytic findings, participants who received a warning communicating both the ability to identify faking and the consequences of faking received significantly lower scores on various faking indices as well as substantive personality measures.

These results are consistent with what VIE theory (Vroom, 1964) would predict. VIE theory asserts that there are three primary drivers of behavior: expectancy, instrumentality, and valence. Expectancy is the belief that applying effort will allow a behavior to be performed, instrumentality is an individual's belief that their behavior will be associated with a particular outcome, and valence is the value the individual places on the outcome. Warning participants to not fake their responses to survey items, especially if this warning communicates the ability of the test administrators to identify faking,

should reduce participants' expectancy, or their belief that they have the ability to fake successfully (Ellingson & McFarland, 2011). Warnings that communicate a consequence also provide an instrumentality associated with faking by asserting that faking behavior will result in an undesirable outcome. This should therefore increase participants' motivation to respond to survey items honestly.

In an IER context, the participants' expectancy is the belief that applying effort will allow them to complete the survey successfully. Although a survey respondent's expectancy depends on a number of factors, such as available resources and how long or demanding the survey is, the vast majority of respondents should believe that they have the ability to complete the survey. The other two motivational influences—instrumentality and valence—are more important in this context and can provide a justification for why warnings should effectively reduce IER. A large proportion of survey respondents in psychological research are college students participating for course credit (Peterson, 2001), which they receive regardless of how much effort they apply. Therefore, although there may be an instrumentality and valence associated with participating in the survey, there are no considerable motivational influences associated with responding attentively. Warning participants that there will be consequences for failing to complete the survey attentively provides this lacking instrumentality, which should therefore increase motivation. Furthermore, if these consequences include the loss of course credits, as in the Huang et al. (2012) study, the valence is likely to be high because the participants' reason for completing the survey in the first place is to obtain course credits.

Therefore, based on the principles of VIE theory and the findings of studies that have examined warnings in the context of faking, it is not surprising that Huang et al.'s (2012) and Meade and Craig's (2012) results contradict one another. Huang et al.'s (2012) warning communicated both the ability to identify IER as well as the consequences of engaging in IER, increasing participants' motivation to apply effort to the survey. Meade and Craig's (2012) warning communicated neither piece of information, and therefore likely had little effect on participant motivation.

Another popular theory that provides support to the assertion that warnings should reduce IER is situational strength. Situational strength suggests that the qualities of some situations can mask the expression of individual differences, thus reducing variability in behavior (e.g., Chatman, 1989; Forehand & von Haller Gilmer, 1964; Meyer, Dalal, & Hermida, 2010). For example, in a weak situation such as a laid-back social gathering with close friends, extraversion would be highly predictive of talkativeness. However, a strong situation such as taking a standardized test would reduce variability in talkativeness, or the expression of extraversion. Meyer et al. (2010) consider consequences to to be an important component of situational strength; the greater the consequences associated with engaging in a particular behavior, the stronger the situation. Warning participants that there will be consequences to engaging in IER creates a strong situation, which should reduce both within-person and between-person variability in IER.

Benefits of IER Prevention. There are many benefits associated with IER prevention that warrant investigating the efficacy of prevention methods. All of these benefits stem from the fact that the ability to prevent IER would obviate the use of the

IER detection methods, which are plagued by problems. Firstly, embedded IER detection scales take up limited survey space that can be conserved if IER prevention is utilized instead. Although another solution to this problem would be to employ the post-hoc IER detection indices, they are not without their own disadvantages, one of which is that they are labor-intensive to compute. Many require the researcher to develop their own syntax or employ cumbersome procedures, which is made even more difficult by the pronounced lack of accessible information regarding how to calculate these indices. Another problem with the post-hoc IER detection indices that prevention circumvents is that they are typically most effective at identifying *either* the random or nonrandom IER response pattern. This necessitates the use of multiple indices—which again are inconvenient and labor-intensive to compute—in order to ensure the detection of an adequate number of inattentive respondents (Meade & Craig, 2012).

Another serious problem inherent to all of the IER detection scales and indices is that they require researchers to establish appropriate cutoff criteria designating which participants responded with sufficient carelessness to be excluded from subsequent analyses. This requires making a subjective judgment call in which the researcher must carefully consider the number of potentially attentive participants who will be flagged and the number of potentially inattentive participants who will not be flagged by a particular cutoff. Unfortunately, because it is more critical to not omit data provided by attentive participants, researchers typically choose to set a very conservative cutoff that is unlikely to exclude even a moderate number of inattentive participants. If IER prevention is effective at deterring most participants from engaging in IER who otherwise would have, this entire process can be averted. Ideally, IER prevention will result in

fewer attentive participants and a greater number of inattentive participants being excluded from subsequent analyses than if IER detection methods were used and the aforementioned process was adhered to. In other words, it is quite possible that IER prevention will result in higher quality data than what would be obtained by using the IER detection methods to identify only the most egregiously careless participants.

Even if the issue of subjective cutoff criteria is disregarded, the fact that the IER detection methods require data to be removed is problematic in its own right. The most obvious issue associated with discarding data is that statistical power is reduced. If this is something that the researcher anticipates, he or she can collect data from a greater number of participants than necessary in order to ensure acceptable statistical power even after cleaning the data for IER. Unfortunately, because participants are a limited resource this is not always feasible, and the researcher must accept an unsatisfactory level of statistical power.

Another less obvious issue that discarding data presents is that there is a chance it may introduce systematic bias. Researchers have demonstrated that personality variables are correlated with willingness to participate in research (e.g., Rogelberg & Luong, 1998; Rosenthal & Rosnow, 1975), and IER can be one manifestation of an unwillingness to participate in research. Therefore, it is highly likely that IER is correlated with substantive variables that psychologists are interested in measuring, such as conscientiousness, counterproductive work behavior (CWB), burnout, or fatigue. To the extent that IER is correlated with the substantive variables of interest, bias will result from discarding data provided by inattentive respondents. Supporting this notion, Bowden (1986) found that when willingness to respond to surveys was correlated with

the substantive variables in the survey, nonresponse—which results in a loss of data much as if the data were removed by the researcher—affected the results obtained from correlational and multivariate statistical analyses. Because IER prevention does not require the removal of data, one principal advantage of this method is that researchers do not have to be concerned with these issues.

There are clearly many practical benefits associated with preventing IER. As such, more research is needed to determine whether IER prevention is effective enough to warrant the recommendation of this approach over IER detection. The results of studies examining the efficacy of warnings to not fake on personality assessments suggest that a warning communicating both the consequences associated with IER and the ability of the test administrators to identify IER is likely to be effective. Therefore, my research design will include a control condition and an experimental condition employing such a warning, which will allow me to evaluate the following hypothesis:

> *Hypothesis 1:* Scores on the IER detection measures will indicate a lower prevalence of IER in the warning condition compared to the control condition.

Additionally, I will provide further evidence for the utility of IER prevention by examining the effects of the warning manipulation on several indicators of data quality. First, I will use Bowling et al.'s (In preparation) method of measuring context effects to test the following hypothesis:

> *Hypothesis 2:* Context effects will occur more frequently in the warning condition compared to the control condition.

The research design employed by this study will also allow me to examine the effects of IER and warnings on internal consistency reliability coefficients and interconstruct correlations. Participants in the control condition responded to the survey under normal conditions. Therefore, I will examine the effects of IER on internal consistency reliability and interconstruct correlations within the control condition. I will also be able to assess the effects of warnings on internal consistency reliability and interconstruct correlations by comparing the estimates obtained from the warning condition with the estimates obtained from the control condition. Because of the difficulty in predicting the specific effects of IER on internal consistency reliability and interconstruct correlations, I propose the following research questions:

Research Question 1: What effect does IER have on the internal consistency reliability of substantive measures?

Research Question 2: What effect does a warning manipulation have on the internal consistency reliability of substantive measures?

Research Question 3: What effect does IER have on the observed relationship between substantive variables?

Research Question 4: What effect does a warning manipulation have on the observed relationship between substantive variables?

If my results demonstrate that the warning manipulation substantially reduces the prevalence of IER, results in a greater manifestation of context effects, and produces internal consistency reliability coefficients and interconstruct correlations coefficients

that are significantly different from those observed in a control condition, this will suggest that researchers should consider using warnings in order to avoid the aforementioned disadvantages of using the IER detection methods.

II. METHOD

Participants

The sample comprised undergraduate students drawn from a participant pool primarily composed of students enrolled in introductory psychology courses at a medium-sized Midwestern university. I excluded 14 subjects from the analyses due to excessive missing data or repeated testing attempts, resulting in a final *N* of 287. The average participant was 19.54 years old; 58% of participants were female; 67% were Caucasian.

Experimental Design

Participants were randomly assigned to either a control condition or to a warning condition. Participants in the warning condition were warned at the outset of the study not to engage in IER.

Procedure

Participants who signed up for the study were provided with an online link to the survey, which was administered through Qualtrics. A feature of Qualtrics was used to randomly assign participants to one of the two survey conditions.

Manipulations

There were 148 respondents in the control condition and 139 respondents in the warning condition. Participants in the warning condition received the following warning prior to beginning the survey:

> "It is vital to our study that participants devote their full attention to this survey. Otherwise years of effort (the researchers' time and the time of other participants) could be wasted. Please be aware that at the end of this

> survey, we will ask you to complete a multiple-choice quiz. This quiz will assess your knowledge of the content of the questionnaire and will be used to determine whether you have been paying attention. If you do not pass this quiz, you might not receive course credit for completing the survey."

In order to ensure that participants in the control condition were sufficiently motivated to allocate effort to the ICRA quiz, they received the following warning after completing the survey, but prior to taking the ICRA quiz:

> "It is vital to our study that participants devote their full attention to this survey. Otherwise, years of effort (the researchers' time and the time of other participants) could be wasted. You will now complete a multiple-choice quiz. This quiz will assess your knowledge of the content of the questionnaire and will be used to determine whether you have been paying attention. If you do not pass this quiz, you might not receive course credit for completing the survey."

All participants were awarded course credit for participating in the study regardless of whether they passed the ICRA quiz. Participants were shown the following debriefing statement on completion of the survey:

> "Thank you for your participation in this survey. You will receive course credit through SONA as compensation for your time and effort regardless of whether or not you passed the multiple-choice quiz. The true purpose of this study was to assess the effectiveness of warning participants as a way to prevent careless and inattentive responses to survey questions.

> Therefore, it was necessary to lead you to believe that there would be consequences for failing to attend to the survey and put forth sufficient effort. We ask that you please not share the true purpose of this study or the content of the survey questions with anyone you know that may participate in the survey. This study has very important implications to the way research is conducted, and it is therefore critical that our results are not biased by participants having prior knowledge of the study methods. Again, thank you for your time and participation."

Manipulation Check

To ensure that participants actually read the warning message, the questions "How will the level of attention you give to this survey be assessed?" and "What are the possible consequences of not giving your full attention to this survey?" were presented below the warning message in both survey conditions. Participants were unable to proceed to the next page until they answered these questions by typing into text boxes.

Measures

Two objectives determined the construction of the survey for this study. First, the survey needed to be relatively long in order to produce a sufficient amount of IER (Berry et al., 1992). Second, it was necessary to include predictor and criterion variables relevant to college students in order to investigate Research Questions 1 through 4. Table 1 illustrates the relationships that I would expect to find amongst these variables based on prior research. Only variable combinations for which the causal direction was clear were used to investigate Research Questions 3 and 4, because moderator effects are dependent

on the direction of causation between the predictor variable and the criterion variable (Judd & Kenny, 2010).

Substantive Predictor measures.

International Personality Item Pool. My primary predictor measure was the 100-item version of the International Personality Item Pool (IPIP; Goldberg, 1992). This measure was designed to correlate highly with the Five Factor Model of personality, and includes 20 items measuring each factor. The survey asks participants how accurately each item describes them, and was administered on a 7-point graphic rating scale from 1 (*very inaccurate*) to 7 (*very accurate*). Example items include "Talk to a lot of different people at parties," which measures Extraversion; "Sympathize with others' feelings," which measures Agreeableness; "Leave my belongings around," which measures Conscientiousness; "Get stressed out easily," which measures Neuroticism; and "Have a vivid imagination," which measures Openness to Experience (see Appendix A for a complete list of the items). The scale was computed by recoding the reverse-keyed items and summing the item scores for each factor. The factors were scored such that higher scores indicated a participant possessed a higher level of the trait. Reliabilities obtained from the full sample in this study were .93 for Extraversion, .90 for Agreeableness, .90 for Conscientiousness, .93 for Neuroticism, and .89 for Openness to Experience.

Negative and Positive Affectivity. Trait negative and positive affectivity were also included as predictor measures, and were assessed by the PANAS (Watson, Clark, & Tellegen, 1988). The PANAS asks participants to indicate how often they generally experience ten positive affective states (e.g., "excited"), and ten negative affective states (e.g., "guilty"; see Appendix B for a complete list of the items). The scales were

administered on a 7-point graphic rating scale from 1 (*never*) to 7 (*always*). The scales were computed by summing the item scores, with higher scores indicating that a participant possessed higher levels of the trait. Reliabilities obtained from the full sample in this study were .88 for positive affectivity and .87 for negative affectivity.

Substantive Criterion Measures.

Family Social Support. Family social support was assessed with a modified version of Abbey, Abramis, and Caplan's (1985) Positive Social Support scale, which includes five items that were administered on a 7-point graphic rating scale from 1 (*never*) to 7 (*always*). One item from Abbey et al.'s (1985) original 6-item measure ("How much have people in your life cared about you as a person") was not included in this study because it decreased the internal consistency reliability of the measure during pilot testing. The items were altered to specifically reflect social support provided by family members (e.g., "How much has your family helped out when too much needed to get done?"; see Appendix C for a complete list of the items). The scale was computed by summing the item scores, with higher scores indicating that a participant experienced greater levels family support. The Cronbach's alpha obtained from the full sample in this study was .93.

Life Satisfaction. Life Satisfaction was assessed with the Satisfaction with Life Scale (SWLS; Diener, Emmons, & Larsen, 1985). It includes five items (e.g., "I am satisfied with my life") administered on a 7-point graphic rating scale from 1 (*strongly disagree*) to 7 (*strongly agree*; see Appendix D for a complete list of the items). The scale was computed by recoding the reverse-keyed items and summing the item scores,

with higher scores indicating that a participant experienced greater life satisfaction. The Cronbach's alpha obtained from the full sample in this study was .85.

General Health. General Health was assessed with the 12-item version of the Physical Symptoms Inventory (PSI-12; Spector & Jex, 1998). It asks how often participants have experienced 12 different physical symptoms (e.g., "an upset stomach or nausea") over the past month (see Appendix E for a complete list of the items). The items were administered on a 5-point graphic rating scale from 1 (*not at all*) to 5 (*every day*). The scale was computed by summing the item scores, with higher scores indicating that a participant experienced more physical symptoms. The Cronbach's alpha obtained from the full sample in this study was .85.

Embedded IER Detection Scales.

Bogus Item Scale/IERS. A pilot study demonstrated that the correlation between Meade and Craig's (2012) Bogus Item Scale and Huang et al.'s (In press) IERS was high enough ($r = .90$, $p < .01$) to justify combining the two scales into one overall measure of IER. After excluding the more outlandish items (e.g., "I am paid biweekly by leprechauns"; Meade & Craig, 2012), I chose to include nine of the remaining items—six items from the Bogus Item Scale and three items from the IERS—that resulted in the highest Cronbach's alpha in the pilot study. An example Bogus item that was included is "I have been to every country in the world" and an example IERS item that was included is "I can run two miles in two minutes" (see Appendix F for a complete list of the items). The scale was administered on a 7-point graphic rating scale from 1 (*very innacurate*) to 7 (*very accurate*). A response of very inaccurate or inaccurate was a correct response to seven of the items, and a response of very accurate or accurate was a correct response to

two of the items. The scale was computed by recoding the items so that a 1 represented an incorrect response and a 0 represented a correct response. The item scores were then summed, with higher scores indicating a greater likelihood that a participant had engaged in IER. The Cronbach's alpha obtained from the full sample in this study was .89.

Self-Reported Diligence Scale. Meade and Craig's (2012) SR Diligence scale was included as a measure of self-reported study engagement. The scale is comprised of eight items that ask participants how much effort they put into the survey (e.g., "I carefully read every item"; see Appendix G for a complete list of the items). The items were administered on a 7-point graphic rating scale from 1 (*strongly disagree*) to 7 (*strongly agree*). In order to be consistent with the other IER measures included in this study, the scale was computed by recoding the positively-keyed items and then summing the item scores so that higher scores indicated that a participant put less effort into their responses. The Cronbach's alpha obtained from the full sample in this study was .90.

Instructed response items. A scale composed of three instructed response items (e.g., "Please select strongly disagree for this question.") were included in this study (see Appendix H for a complete list of the items). Two of the items were administered on a 7-point graphic rating scale from 1 (*very inaccurate*) to 7 (*very accurate*), and one item was administered on a 7-point graphic rating scale from 1 (*strongly disagree*) to 7 (*strongly agree*). The scale was computed by recoding the items so that a 1 represented an incorrect response and a 0 represented a correct response. The item scores were then summed, with higher scores indicating a greater likelihood that a participant engaged in IER. The Cronbach's alpha obtained from the full sample in this study was .77.

ICRA. The Item Content Recognition Approach (ICRA; Bowling et al., In preparation) scale consists of ten target items that are distributed throughout the survey and administered on a 7-point graphic rating scale from 1 (*very inaccurate*) to 7 (*very accurate*). These items are not scored, and are only included to provide memorable items (e.g., "If my friends dared me to eat a live goldfish, I would probably do it") over which participants were quizzed at the conclusion of the survey. The quiz is composed of ten multiple-choice items (e.g., "Earlier in this questionnaire, we asked you about eating _____ as part of a dare") that have four response options each (see Appendix I for a complete list of the items). The scale was computed by recoding the quiz items so that a 0 represented a correct response and a 1 represented an incorrect response. The item scores were then summed, with higher scores indicating a greater likelihood that a participant was engaging in IER. The Cronbach's alpha obtained from the full sample in this study was .89.

Post-Hoc IERS Detection Indices.

Psychometric antonyms and psychometric synonyms. Item pairs with correlations of -.60 or stronger and item pairs with correlations of .60 or stronger were identified and constituted the psychometric antonym and synonym indices respectively (as described in Meade & Craig, 2012). The psychometric synonym index was composed of 53 such item pairs, and the psychometric antonym index was composed of 8 such item pairs. Scores were obtained for each participant by computing the within-person correlations between each set of items. The correlation for the psychometric synonyms was then reversed to produce a score where higher values were more indicative of IER.

Individual reliability. Individual reliability (as described in Jackson, 1976) was computed by splitting the 10 substantive unidimensional scales into an even subscale and an odd subscale based on the order of item presentation. The correlation between the even and odd subscales was then computed for each participant, and the correlation was reversed to produce scores where higher values are more indicative of IER.

Long string indices. Visual Basic syntax was formulated to calculate the maximum long string and average long string indices. The score for the maximum long string index was simply the maximum number of times that the same response option was utilized on consecutive items, and the score for the average long string index was the average of each long string of responses across all of the survey pages. Therefore, higher scores on each of these indices were more indicative of IER.

Response time. Survey time, or the amount of time that a participant spent on the entire survey, was used to represent response time. Very short survey times were considered indicative of IER. Participants' times were recoded such that higher scores were more indicative of IER. Some participants' survey times were excessively long in duration, which suggests that they took a break from the survey and returned to complete it at a later time. Descriptive statistics revealed that 97% of the sample was able to complete the survey in less than 170 minutes. Therefore, following the procedure used by Meade and Craig (2012), survey times greater than 170 minutes were set as missing for eight participants.

IER Consequences Measure

Context effects. Context effects were measured by the inclusion of one item, administered with a dichotomous true/false response format, that was designed to prime

participants to respond to a later, open-ended item with the same idea. Participants were randomly assigned to one of two priming items: "I would like to visit London, England someday," or "I would like to visit Rome, Italy someday," and then all participants received the open-ended item "What are three common international travel destinations?" on the next page of the survey. The open-ended item was scored as 9 (not answered), 0 (context effect absent), or 1 (context effect present).

Survey Arrangement

Survey items were spread across 15 web pages. The study cover letter was presented on page 1; the warning condition received the warning message and 2 manipulation check items on page 2, whereas the control condition received an alternative "click '>>' to begin the survey" message on page 2; pages 3 through 10 contained the substantive measures (IPIP, PANAS, family social support, life satisfaction, physical symptoms) with the embedded IER detection items (ICRA target items, Bogus items, instructed response items, repeated items) distributed throughout; page 11 contained the context effect priming item; page 12 contained the open-ended context effect item, three demographic items (age, gender, and race), and two academic performance items (current GPA and expected GPA); page 13 contained the nine Diligence items and the UseMe item; the control condition received the warning message and 2 manipulation check items on page 14, whereas the warning condition received an alternative "click '>>' to begin the multiple-choice quiz" message on page 14; and page 15 contained the 11 ICRA multiple-choice items. See Table 2 for a summary of the survey arrangement for each condition.

III. RESULTS

Manipulation Check

Participants' responses to the open-ended manipulation check items were coded according to whether or not they reflected an understanding of the warning message. 76.3% of participants' responses to the first manipulation check item ("How will the level of attention you give to this survey be assessed?"), and 90.9% of participants' responses to the second manipulation check item ("What are the possible consequences of not giving your full attention to this survey?") clearly indicated that they had attended to the warning message. The majority of participants whose responses did not reflect knowledge of the warning message's content appeared to have misunderstood what the items were asking.

Descriptive Statistics

The means, standard deviations, internal consistency reliabilities, and correlations calculated for the IER detection indices and the substantive variables are displayed in Tables 3 and 4 respectively. With a few exceptions, the IER measures exhibit moderate to high correlations with one another, supporting the construct validity of these indices. One notable finding is that survey time does not correlate highly with the other IER indices, which suggests that the use of response time as a measure of IER in this study is problematic. Another interesting finding is that scores on the psychometric synonym and individual reliability indices were not significantly related to scores on the long string indices, which supports the idea that these measure different types of IER response styles. However, scores on the psychometric antonym index, the other index in this study that takes the inconsistency approach to IER detection, were significantly related to scores on

the long string indices. This may have occurred because some of the substantive scales were composed solely of items scored in one direction. Responses to many of these items were highly correlated and were included in the psychometric synonym index, and these scales were also included in the individual reliability index. Therefore, although participants engaging in the long string IER response style on these scales would have scores indicating probable IER on the long string indices, their scores on the psychometric synonym and individual reliability indices would reflect conscientious responding. This could have attenuated the correlation between these IER detection indices within the overall sample.

Tests of Hypotheses

Hypothesis 1: Efficacy of Warnings. I found support for the hypothesis that warning participants to attend to the survey would reduce the prevalence of IER. Ten independent samples *t*-tests were conducted, and the Bonferroni procedure was applied to correct for family-wise error. The results of these *t*-tests indicated that there were significant differences between the warning condition and the control condition on four out of the 10 IER detection indices (see table 5). Scores on the ICRA ($t(285) = 3.56$, $p < .005$, $d = 0.42$), bogus items ($t(285) = 4.57$, $p < .005$, $d = 0.53$), instructed response items ($t(285) = 4.43$, $p < .005$, $d = 0.51$), and self-reported diligence scale ($t(285) = 3.81$, $p < .005$, $d = 0.45$) reflected significantly less IER in the warning condition compared to the control condition.

Hypothesis 2: Context Effects. Participants' responses to the open-ended context effect item were coded according to whether or not they mentioned the international travel destination that was the subject of the context effect priming item.

Logistic regression was performed for each context effect condition to investigate the effects of the context manipulation on participants' coded responses to the open-ended context effect item. These analyses revealed that participants who were primed with "London" were significantly more likely to mention London in their response to the open-ended item ($W = 15.41$, $p < .001$), and participants who were primed with "Rome" were significantly more likely to mention Rome in their response to the open-ended item ($W = 10.26$, $p < .001$). Additional logistic regression analyses were conducted using the full sample without regard to the context manipulation in order to test the hypothesis that warning participants to attend to the survey would result in a greater incidence of context effects, and to determine whether the IER detection indices were significant predictors of the occurrence of context effects. These analyses indicated that there were no statistically significant differences in the occurrence of context effects between experimental conditions ($W = .04$, $p = .84$), and that none of the IER indices were statistically significant predictors of the occurrence of context effects (see Table 6). Thus, hypothesis 2 was not supported.

Research Question 1: Effects of IER on Internal Consistency Reliability.

Research Question 1 concerned the effects of IER on the internal consistency reliability estimates of substantive measures. Participants were split into two comparison groups for each IER measure; one group was composed of the participants who achieved the lowest 75% of scores on the IER measure (i.e., more attentive respondents), and the other group was composed of the participants who received the highest 25% of scores on the IER measure (i.e., less attentive respondents). Cronbach's alphas for each substantive scale were computed separately for each group, and the estimates of the two groups for

each IER measure were compared. The control condition was used for this analysis, because the incidence of IER within the control condition should be comparable to the incidence of IER within a typical sample. Therefore, using the control condition to evaluate Cronbach's alphas is more realistic than using the full sample due to the reduced variability of IER within the warning condition.

Differences in Cronbach's alphas between the comparison groups ranged from -0.37 to 0.14 (see Table 7). The Feldt test was used to test the statistical significance of each of these differences, and the set of analyses conducted for each IER measure was regarded as one analysis when correcting for family-wise error with the Bonferroni procedure. Based on Huang et al.'s (2012) observation that IER can inflate internal consistency reliability estimates for scales that are primarily scored in one direction, I expected the Cronbach's alphas of PA, NA, Family Social Support, Life Satisfaction, and Physical Symptoms, which are scored in one direction, to be higher among the participants who received the highest 25% of scores on the IER measures. Alternatively, I expected the Cronbach's alphas of the Big Five measures, which have reverse-scored items, to be lower among the participants who received the highest 25% of scores on the IER measures. 62% of the significant differences in Cronbach's alphas for the scales scored in one direction were higher among participants who received the highest 25% of scores on the IER measures as expected, and 100% of the significant differences in Cronbach's alphas for the Big Five measures were lower among participants who received the highest 25% of scores on the IER measures as expected. This pattern of results largely supports Huang et al.'s (2012) prediction.

Table 8 displays the percentage of differences in Cronbach's alphas that were significantly higher among the participants who received the highest 25% of scores on the IER measures, the percentage of differences in Cronbach's alphas that were significantly lower among the participants who received the highest 25% of scores on the IER measures, and the percentage of differences in Cronbach's alphas that were nonsignificant. There is one especially notable finding within this table: all of the statistically significant differences in the Cronbach's alphas of the measures scored entirely in one direction between the comparison groups on the psychometric synonym and individual reliability indices were in the opposite direction of what was expected based on Huang et al.'s (2012) observation. One likely explanation for this finding will be offered in the discussion section.

Research Question 2: Effect of Warning on Internal Consistency Reliability. Research Question 2 concerned the effects of a warning manipulation on the internal consistency reliability estimates of substantive measures. Cronbach's alphas for each substantive scale within the control condition were compared to Cronbach's alphas for each substantive scale within the warning condition. Differences in Cronbach's alphas between conditions ranged from -0.03 to 0.07 (see Table 9). The Feldt test was used to test the statistical significance of these differences. The differences between the conditions' Cronbach's alphas for conscientiousness ($W = .75$, $p < .04$) and physical symptoms ($W = .65$, $p < .01$) were statistically significant, although only the difference between the conditions' Cronbach's alphas for physical symptoms remained significant after correcting for family-wise error using the Bonferroni procedure.

Research Question 3: Effect of IER on Interconstruct Correlations. Research Question 3 concerned the effects of IER on interconstruct correlations. Regression analyses were conducted for the fifteen predictor-criterion relationships depicted in table 1, using each IER index as a moderator of each relationship. The set of regression analyses conducted for each IER measure was regarded as one analysis when applying the Bonferroni procedure to correct for family-wise error. Again, the control condition was used for these analyses as opposed to the full sample, because the full sample is less representative of a typical sample that a researcher might encounter due to the reduced variability of IER within the warning condition. Table 10 displays the results of these analyses across substantive predictor-criterion relationships, and Table 11 displays the results of these analyses across IER indices. 19 of the 150 moderator effects that were examined were statistically significant, and only three of these remained significant after correcting for family-wise error. Further investigation revealed that the relationship between positive affectivity and life satisfaction was inflated by participants who performed poorly on the instructed response items, the relationship between negative affectivity and life satisfaction was inflated by participants who performed poorly on the individual reliability index, and the relationship between negative affectivity and life satisfaction was attenuated by participants who performed poorly on the psychometric synonym index.

Research Question 4: Effect of Warning on Interconstruct Correlations. Research Question 4 concerned the effects of a warning manipulation on interconstruct correlations. Regression analyses were conducted for the fifteen predictor-criterion relationships depicted in Table 1, using condition as a moderator of each relationship.

The results of these analyses are displayed in Table 12. The ΔR^2 obtained when adding condition as a predictor ranged from .000 to .011, none of which were statistically significant.

IV. DISCUSSION

The current study provides some evidence in support of previous research findings suggesting that IER has consequences to the quality of self-report data. In addition, this study demonstrated the value of implementing warnings as an IER prevention strategy. I found support for one of my two hypotheses and gained insights into the four research questions that I posed. The subsequent discussion offers possible explanations for the findings of this study as well as the implications these findings may have on future applications and research involving IER.

Value of IER Prevention

The current study demonstrated a lower prevalence of IER in the warning condition as compared to the control condition, supporting Hypothesis 1. The mean differences in scores between experimental conditions were in the hypothesized direction for all 10 of the IER detection indices. Four of these mean differences reached statistical significance and had moderate effect sizes, which are similar to the effect sizes observed in studies documenting the effects of warnings on faking (Dwight & Donovan, 2003). This suggests that warnings are an effective IER prevention method.

These results also provide evidence for the construct validity of the four IER detection methods that identified significantly fewer inattentive participants in the warning condition compared to the control condition. Because we would expect fewer participants to respond inattentively after being warned, the fact that these results were actually observed suggests that the embedded detection scales and self-reported Diligence scale are in fact identifying inattentive respondents. These indices might have yielded stronger support for Hypothesis 1 because they have the ability to detect both random and

nonrandom IER response patterns and thus, are more powerful IER detection tools. The psychometric synonym, psychometric antonym, and individual reliability indices are primarily measures of random IER response patterns, whereas the long string indices are primarily measures of nonrandom IER response patterns. The nonsignificant results for survey time can be explained by the low correlations between this measure and the other IER detection indices, which is evidence that survey time was not an effective measure of IER in this study.

IER and Context Effects

The tests of Hypothesis 2 failed to demonstrate a relationship between the incidence of context effects and experimental condition or scores on the IER detection indices. This could indicate that IER does not reduce the occurrence of context effects as predicted. An alternative explanation for this finding is that other factors influenced the manifestation of context effects and obscured the effects of IER. For example, one reason context effects may not have occurred among some attentive participants is that the participants gave an honest, thoughtful response to the open-ended context effect item, which was not represented in the priming context effect item. In addition, although context effects are unlikely to occur amongst the most negligent participants, they may actually be more likely to occur amongst moderately inattentive participants that are allocating a sufficient amount of effort to read and respond to the survey items, but an insufficient amount of effort to provide true or thoughtful responses. These participants may take advantage of mental shortcuts that are available in the survey, and thus simply provide the same response to the open-ended context effect item that was given in the priming context effect item. This suggests that the method used to measure the influence

of IER on context effects in this study may have been ineffective, and moreover, that the influence of IER on context effects may be difficult to measure in general.

IER and Internal Consistency Reliability

The examination of Cronbach's alphas for Research Question 2 largely supports Huang et al.'s (2012) results demonstrating that IER attenuates internal consistency reliability estimates for scales that are scored in both directions and inflates internal consistency reliability estimates for scales that are primarily scored in one direction. All of the significant differences in Cronbach's alphas for the measures with reverse-scored items were lower among participants who were more likely to have engaged in IER, and the majority of the significant differences in Cronbach's alphas for the substantive measures scored in one direction were higher among participants who were more likely to have engaged in IER. The relative prevalence of each IER response pattern may provide insight into the finding that more of the significant differences in Cronbach's alphas for the measures with reverse-scored items were in the anticipated direction. As Huang et al. (2012) explained, inattentive participants who repeatedly select the same response option (nonrandom IER response patterns) on scales that are primarily worded in one direction should serve to inflate internal consistency reliability estimates. However, I would emphasize that random responses on these scales should still attenuate internal consistency reliability estimates, and it is the combination of these two effects that determines the ultimate impact of IER. Therefore, the extent to which Cronbach's alphas on scales that are primarily worded in one direction are attenuated or inflated depends on the percentage of inattentive respondents in the sample that are engaging in

each type of IER response pattern. This suggests that it may be difficult to predict the effects of IER on the internal consistency reliability of these types of scales.

The results of the evaluations of Cronbach's alphas between the two comparison groups for each IER measure appear to support this conjecture. Most notably, all of the significant differences in the Cronbach's alphas of the substantive measures scored entirely in one direction were lower among participants who the psychometric synonym and individual reliability indices indicated were more likely to have engaged in IER. This is opposite of what would be expected based on Huang et al.'s (2012) observation. However, because these measures primarily detect random IER response patterns, this finding supports the idea that random responses attenuate the internal consistency reliability estimates of substantive measures scored in one direction. The results of this investigation suggest that the way in which IER impacts the internal consistency reliability estimates of scales depends on the relative prevalence of each type of IER response pattern within the sample. Furthermore, these results underscore the importance of using an IER reduction technique, whether detection or prevention, that captures both types of IER response patterns.

Warnings and Internal Consistency Reliability

The results of the analyses performed to investigate Research Question 2 revealed small differences in Cronbach's alphas between experimental conditions, only one of which was statistically significant: the Cronbach's alpha for the physical symptoms scale was higher in the control condition than in the warning condition. One reason this was the only statistically significant difference observed may be due to the position of the physical symptoms scale within the survey; it was the last substantive measure in the

survey and came after participants had already responded to 211 items. Past research has found a positive relationship between IER and survey length (Baer et al., 1997; Berry et al., 1992). Therefore, it is likely that IER among participants in the control condition, who had no incentive to refrain from engaging in IER, increased as the survey progressed. Alternatively, participants in the warning condition did have an incentive to refrain from engaging in IER, and thus likely applied sufficient effort to their responses throughout the duration of the survey. Consequently, I would expect differences in data quality, including internal consistency reliability estimates, between conditions to be greater for substantive measures at the end of the survey compared to substantive measures at the beginning of the survey.

IER and Interconstruct Correlations

Of the 150 moderator effects that were examined as an investigation of Research Question 3, only three were statistically significant after correcting for family-wise error. This suggests that IER does not have a significant impact on interconstruct correlations, either because the incidence of IER in the sample was not of sufficient prevalence to have an effect or because of the complex mechanisms that influence the effects of IER on interconstruct correlations. Another possible explanation for the shortage of significant findings is low statistical power. The control condition was used for these analyses as opposed to the full sample due to the reduced variability of IER within the warning condition. This limited the sample size to 148. A power analysis revealed that the probability of detecting a small effect, which is the expected effect size in tests of moderated regression, was only 0.31.

Warnings and Interconstruct Correlations

The results of the analyses performed to investigate Research Question 4 revealed that experimental condition did not significantly moderate any of the relationships examined. This could suggest that IER does not influence the results of correlational analyses, which is consistent with the previously discussed findings in relation to Research Question 3. Alternatively, these results could reflect the inability of the manipulation to completely eradicate IER. The manipulation will not be effective on participants who do not read the warning message or participants who are suspicious of the warning and want to test its validity. Therefore, it is possible that there were no differences between experimental conditions because the IER that was engaged in by these types of participants impacted the interconstruct correlations observed in the warning condition, causing them to be similar to the interconstruct correlations observed in the control condition. Another possible explanation for the lack of significant effects is low statistical power. A power analysis revealed that the probability of detecting a small effect, which is the expected effect size in tests of moderated regression, was 0.56.

Implications

This study has many implications for the way researchers and practitioners administer surveys and analyze data. My results suggest that warnings are an effective way to reduce the prevalence of IER and provide evidence for the construct validity of the IER detection indices. As expected, the mean differences of scores on all of the IER detection indices between experimental conditions were in the hypothesized direction, indicating a lower prevalence of IER in the warning condition. This supports the claim that these indices detect IER.

My results also support Huang et al.'s (2012) findings demonstrating the specific ways in which IER influences internal consistency reliability estimates. IER appears to attenuate internal consistency reliability estimates for scales that are scored in both directions and inflate internal consistency reliability estimates for scales that are primarily scored in one direction. Furthermore, I found that warnings can affect internal consistency reliability estimates, presumably through a reduction in IER. Although I did not find much evidence to support the claim that IER or warnings impact observed relationships between substantive variables, the low statistical power of the analyses performed to investigate these phenomena preclude drawing conclusions from these results. In addition, past research has demonstrated that IER does affect observed interconstruct correlations (Huang et al., In press; Huang et al., In review). Altogether, these findings imply that addressing IER is an important consideration when designing a research study in order to maximize the quality of the resulting data.

Researchers interested in addressing IER in their studies can either use one or more of the various IER detection measures or an IER prevention method; however, there are many advantages associated with IER prevention. One of the benefits of IER prevention is that it is not necessary to include extra measures that take up precious survey space and increase the burden placed on participants, or to compute labor-intensive post-hoc indices. Similarly, the effectiveness of IER prevention does not depend on the type of IER response pattern, the frequency of inattentive responses, or the distribution of IER data. Another advantage of using IER prevention over detection is that researchers do not have to discard data and can retain the full statistical power of their sample. Furthermore, because it is highly likely that IER is correlated with

constructs that psychologists frequently study, such as conscientiousness, counterproductive behavior, burnout, or fatigue, discarding data provided by inattentive respondents may result in a loss of variability or have biasing effects. This can be avoided by implementing IER prevention methods. Due to these advantages and the observed reduction of IER in the warning condition, my study supports the use of IER prevention methods in lieu of detection and removal practices.

The only IER prevention methods that have been empirically examined thus far are warnings (Huang et al., 2012; Meade & Craig, 2012) and identity disclosure (Meade & Craig, 2012). The effectiveness of warnings has received inconsistent support in the IER literature; however, this can be explained by the incomparable warning messages that were used in these studies. Huang et al. (2012) found their warning, which communicated both the ability to identify IER as well as the consequences of engaging in IER, to be an effective IER prevention method. Alternatively, Meade and Craig's (2012) warning, which communicated neither piece of information, did not effectively reduce the prevalence of IER. Based on past studies examining different types of warnings in other contexts, I conjectured that the different information conveyed in the warning messages used by these two studies was responsible for the discrepant findings. In this study, I used a warning similar to the one used by Huang et al. (2012) and found that it effectively reduced the prevalence of IER. This supports my explanation for the contradictory findings of Huang et al.'s (2012) study and Meade and Craig's (2012) study, and is consistent with the finding in the faking literature suggesting that warnings are more effective when they convey both the ability of the test administrators to detect faking and the consequence of faking (Dwight & Donovan, 2003). Therefore, until

additional IER prevention methods have been investigated, I recommend that researchers interested in implementing IER prevention methods in their own studies use warning messages that include these two key pieces of information.

Limitations

One limitation of this study is that a very long survey had to be used in order to achieve sufficient variability in IER, limiting generalizability. However, researchers frequently administer lengthy surveys in order to collect enough data to get multiple publications out of one data collection endeavor. In addition, some participants appear to engage in IER throughout the entire survey regardless of length. It is probably the case that very few surveys are brief enough to completely eliminate the risk of IER.

Another limitation of this study is the possibility that the manipulation did not have the intended effect because participants did not read or understand the warning message. Another possibility is that some participants who had already completed the survey informed participants who had not yet completed the survey of the fact that there would be no consequences to engaging in IER despite the content of the warning message. This might have caused some participants in the warning condition to ignore the warning message and engage in IER, which could potentially have washed out some of the true differences between experimental conditions. However, participants' responses to the open-ended manipulation check items suggest that the majority of participants read and understood the instructions. In addition, because a failure of the manipulation for either one of these reasons would only decrease the likelihood of obtaining significant effects, the fact that I obtained statistically significant results is evidence that a failure of the manipulation was not a significant issue in my study.

There is also a limitation of the psychometric antonym and individual reliability indices in this study that should be mentioned. Only eight substantive item pairs had correlations of -0.6 or stronger and could be used in the psychometric antonym index, and the individual reliability index was only composed of ten substantive scales. Goldberg (1985) recommended using 30 item pairs in the psychometric antonym index in order to produce a stable correlation coefficient, and this principle also applies to the individual reliability index. Therefore, psychometric antonyms and individual reliability were not optimally measured in this study, and this may have influenced the results of the analyses involving these indices.

The inadequate statistical power of the moderated regression analyses is another limitation of this study, and it is likely that this limitation did affect the results of these analyses. A power analysis revealed that the probability of detecting a small effect was only 0.31 for the tests investigating Research Question 3 and 0.56 for the tests investigating Research Question 4. This could explain the shortage of statistically significant moderator effects that I observed.

Future Research Directions

Future research should investigate other potential methods of IER prevention and compare the relative effectiveness of these methods. Meade & Craig (2012) found identity disclosure to effectively deter IER. However, due to the inadequacy of the authors' warning manipulation, it is unclear whether identity disclosure or warnings are more effective IER prevention methods. It would also be useful to assess whether any additional benefit is gained by using IER detection methods in conjunction with IER prevention methods.

This study failed to find much evidence in support of past studies that have shown that IER impacts observed relationships between substantive variables. Future studies should attempt to elucidate this by examining the magnitude of IER's effects on data quality and the idiosyncrasies of the situations in which these effects appear. In addition, it would be valuable to continue to investigate the extent to which IER detection and prevention methods mitigate the negative effects of IER on data quality.

Future studies should also investigate the predictors of IER, including individual differences such as conscientiousness, as well as environmental predictors such as distractions. Knowing which variables are correlated with IER would allow researchers to determine whether using IER detection and removal practices will introduce bias or range restriction into a particular study. This knowledge could also have implications for discouraging IER. For example, if future research determines that environmental distractions are a significant predictor of IER, researchers could attempt to exert environmental control over their studies in order to minimize these distractions and thus reduce IER. However, there are obvious challenges associated with attempting to collect valid self-report data from inattentive respondents. Researchers interested in pursuing this topic should consider alternative methods of data collection, such as significant other-reports.

Conclusion

In conclusion, this study evaluated the extent to which warnings are an effective IER prevention method and investigated some of the consequences to data quality that previous studies have attributed to IER. My findings suggest that warnings effectively reduce the prevalence of IER and provide additional evidence supporting the construct

validity of the IER detection methods. The investigations into IER's consequences produced somewhat mixed results in this study. The findings indicated that both IER and warnings can have an impact on the internal consistency reliability of substantive measures, but provided little evidence to suggest that either affect the occurrence of context effects or interconstruct correlations. However, low statistical power was a problem for all of the tests involving interconstruct correlations in this study, and past research has found that IER can significantly affect relationships between substantive variables (Huang et al., In press; Huang et al., In review). Collectively, the results of my study and past research examining the consequences of IER suggest that IER has the potential to impact data quality. Therefore, it is important to address IER to attempt to mitigate this impact. Warnings may be a favorable alternative to IER detection measures as they are relatively convenient to implement and avoid many of the problems associated with the detection measures.

V. REFERENCES

Abbey, A., Abramis, D. J., & Caplan, R. D. (1985). Effects of different sources of social support and social conflict on emotional wellbeing. *Journal of Basic and Applied Social Psychology, 6*(2), 111-129. doi:10.1207/s15324834basp0602_2

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

Beach, D. A. (1989). Identifying the random responder. *Journal of Psychology: Interdisciplinary and Applied, 123,* 101-103.

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larson, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment, 4*(3)*,* 340-345. doi:10.1037/1040-3590.4.3.340

Bowden, R. J. (1986). Self-selection biases in correlational studies based on questionnaires. *Psychometrika, 51*(2), 313–325. doi:10.1007/BF02293987

Bowling, N. A., Burns, G., Huang, J., & Blackmore, C. E. (In preparation). The Item Content Recall Approach: A new strategy for detecting insufficient effort responding to self-report questionnaires. Target journal: *Journal of Applied Psychology*.

Charter, R. A. (1994). Determining random responding for the Category, Speech-Sounds Perception, and Seashore Rhythm tests. *Journal of Clinical and Experimental Neuropsychology, 16,* 744-748.

Chatman, J. A. (1989). Improving Interactional Organizational Research: A Model of Person-Organization Fit. *Academy Of Management Review*, *14*(3), 333-349. doi:10.5465/AMR.1989.4279063

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business Psychology, 25,* 325-334. doi: 10.1007/s10869-010-9181-6

Costa, P. T. Jr., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In D. H. Saklofske (Ed.), *The SAGE handbook of personality theory and assessment. Vol 2: Personality measurement and testing* (pp. 179-198). Thousand Oaks, CA: Sage.

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.

Diener, E. D., Emmons, R. A., & Larsen, R. J. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71-75. doi:10.1207/s15327752 jpa4901_13

Douglas, K. M., & McGarty, G. (2001). Identifiability and self-presentation: Computer-mediated communication and intergroup interaction. *British Journal of Social Psychology, 40,* 399-416.

Dwight, S. A., & Donovan, J. J. (2003). Do Warnings Not to Fake Reduce Faking?. *Human Performance, 16*(1), 1-23.

Ehlers, C., Greene-Shortridge, T. M., Weekley, J. A., & Zajack, M. D. (2009). *The exploration of statistical methods in detecting random responding.* Paper presented at the annual meeting of the Society for Industrial/Organizational Psychology, Atlanta, GA.

Ellingson, J. E. & McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: An application of VIE theory. *Human Performance, 24*(4), 322-337. doi: 10.1080/08959285.2011.597477

Fazio, R. H., & Williams, C. J. (2008). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election. In R. H. Fazio, R. E. Petty (Eds.) , *Attitudes: Their structure, function, and consequences* (pp. 445-459). New York, NY, US: Psychology Press.

Forehand, G. A., & Von Haller, G. (1964). Environmental variation in studies of organizational behavior. *Psychological Bulletin*, *62*(6), 361-382. doi:10.1037/h0045960

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4(1), 26-42. doi:10.1037/1040-3590.4.1.26

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology, 48*, 82-98.

Green, M., DeCourville, N., & Sadava, S. (2012). Positive affect, negative affect, stress, and social support as mediators of the forgiveness-health relationship. *The*

Journal Of Social Psychology, 152(3), 288-307. doi:10.1080/00224545.2011.603767

Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality And Individual Differences, 34*(4), 723-727. doi:10.1016/S0191-8869(02)00057-0

Heller, D., Judge, T. A., & Watson, D. (2002). The confounding role of personality and trait affectivity in the relationship between job and life satisfaction. *Journal Of Organizational Behavior, 23*(7), 815-835. doi:10.1002/job.168

Huang, J.L., Bowling, N.A., Liu, M., & Li, Y. (In press). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology*.

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. (In review). Insufficient effort responding: Uncovering the insidious threat to survey data quality. Second revise and resubmit at *Journal of Applied Psychology*.

Jackson, D. N. (1976). *The appraisal of personal reliability.* Paper presented at the meetings of the Society of Multivariate Experimental Psychology, University Park, PA.

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

Johnson, S. J., Batey, M. M., & Holdsworth, L. L. (2009). Personality and health: The mediating role of trait emotional intelligence and work locus of control. *Personality And Individual Differences, 47*(5), 470-475. doi:10.1016/j.paid.2009.04.025

Judd, C. M. & Kenny, D. A. (2010). Data analysis in social psychology: Recent and recurring issues. In S. T. Fiske, D. T. Gilbert, G. Lindzey (Eds.), *Handbook of social psychology, Vol 1 (5th ed.)* (pp. 115-139). Hoboken, NJ, Us: John Wiley & Sons Inc.

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

Lee, H. (2006). Privacy, publicity, and accountability of self-presentation in an on-line discussion group. *Sociological Inquiry, 76,* 1-22.

Liu, M., Bowling, N. A., Huang, J. L., & Kent, T. A. (2013). Insufficient effort responding to surveys as a threat to validity: The perceptions and practices of SIOP members. *The Industrial-Organizational Psychologist*.

Lumsden (1977). Person reliability. *Applied Psychological Measurement, 1*(4), 477-482. doi:10.1177/014662167700100403

Malhorta, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science, 12,* 1865-1883. doi: 10.1287/mnsc.1060.0597

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

Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management, 36*: 121-140.

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.

Okun, Stock, Haring, & Witter (1984). Health and subjective wellbeing: a meta-analysis. *The International Journal of Aging and Human Development, 19*(2), 111-132. doi: 10.2190/QGJN-0N81-5957-HAQD

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598-609. doi:10.1037/0022-3514.46.3.598

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research, 28*(3), 450-461. doi:10.1086/323732

Rogelberg, S. G., & Luong, A. (1998). Nonresponse to Mailed Surveys: A Review and Guide. *Current Directions In Psychological Science*, (2), 60. doi:10.2307/20182503

Rosenthal, R. & Rosnow, R. L. (1975). *The Volunteer Subject.* New York: Wiley.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: the result of careless respondents? *Applied Psychological Measurement, 9*(4)*,* 367-373. doi:10.1177/014662168500900405

Schuman, H. (1992). Context effects: State of the past/state of the art. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 5-20). New York: Springer-Verlag.

Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context.* New York: Academic Press.

Simsek, Z. & Veiga, J. F. (2001). A primer on internet organizational surveys. *Organizational Research Methods, 4*(3), 218-235. doi:10.1177/109442810143003

Smith, E. & Miller, F. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology, 44*(3), 492-505. doi:10.1037/0022-3514.44.3.492

Spector, P. E. (2002). Research methods in industrial and organizational psychology: Data collection and data analysis with special consideration to professional issues. In N. Anderson, D. S. Ones, H. Sinangil, C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology, Volume 1: Personnel psychology* (pp. 10-26). Thousand Oaks, CA: Sage Publications Ltd.

Spector, P. E., & Jex, S. M. (1998). Development of four self-report measures of job stressors and strain: Interpersonal Conflict at Work Scale, Organizational Constraints Scale, Quantitative Workload Inventory, and Physical Symptoms Inventory. *Journal Of Occupational Health Psychology, 3*(4), 356-367. doi:10.1037/1076-8998.3.4.356

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition 4*(3), 215-230. doi:10.1016/0010-0277(76)90018-4

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers.* San Francisco, CA: Josey-Brass.

Swickert, R. (2009). Personality and social support. In P. Corr & G. Matthews (Eds.), *Cambridge handbook of personality* (pp. 524–540). Cambridge, England: Cambridge University Press.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational And Psychological Measurement, 59*(2), 197-210. doi:10.1177/00131649921969802

Vroom, V. H. (1964). *Work and Motivation.* New York: Wiley.

Watson, D. (1988). Intraindividual and interindividual analyses of positive and negative affect: Their relation to health complaints, perceived stress, and daily activities. *Journal Of Personality And Social Psychology, 54*(6), 1020-1030. doi:10.1037/0022-3514.54.6.1020

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Woods, C. M. (2006). Careless responding to reverse-worded items: implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28*(3)*,* 189-194.

Table 1.
Expected Relationships Between Substantive Variables

	1	2	3	4	5	6	7	8	9	10
1. Extraversion										
2. Agreeableness										
3. Conscientiousness										
4. Neuroticism										
5. Openness										
6. PA										
7. NA										
8. Family Support	$+_1$	$+_1$		$-_1$		$+_2$	$-_2$			
9. Life Satisfaction	$+_3$		$+_3$	$-_3$		$+_4$	$-_4$			
10. General Health	$+_5$		$+_5$	$-_5$		$+_6$	$-_7$			

Note. Swickert (2009)$_1$; Green, DeCourville, & Sadava (2012)$_2$; Hayes & Joseph (2003)$_3$; Heller, Judge, & Watson (2002)$_4$; Johnson, Batey, & Holdsworth (2009)$_5$; Okun, Stock, Haring, & Witter (1984)$_6$; Watson (1988)$_7$.

Table 2.
Summary of Survey Arrangement

Survey Page	Warning Condition (N = 139)	Control Condition (N = 148)
1	Cover Letter	Cover Letter
2	Warning	
2	Manipulation check items	
3-10	Substantive measures with embedded ICRA target items, Bogus items, and Instructed Response Items	Substantive measures with embedded ICRA target items, Bogus items, and Instructed Response Items
11	Context effect priming item	Context effect priming item
12	Context effect open-ended item	Context effect open-ended item
12	Demographic and GPA items	Demographic and GPA items
13	Diligence Scale	Diligence Scale
14		Warning
14		Manipulation check items
15	ICRA multiple-choice items	

Table 3.
Means, standard deviations, internal consistency reliabilities, and correlations of IER indices

	N	M	SD	1	2	3	4	5	6	7	8	9
1. ICRA	287	10.14	2.28	(.89)								
2. Bogus Items	287	8.04	1.99	.75**	(.89)							
3. Instructed Response Items	287	2.69	0.75	.67**	.74**	(.77)						
4. Self-Reported Diligence	287	5.94	1.20	.72**	.51**	.48**	(.90)					
5. Psychometric Antonyms	287	-0.69	0.37	.63**	.55**	.49**	.55**	--				
6. Psychometric Synonyms	287	0.67	0.20	.45**	.51**	.46**	.31**	.33**	--			
7. Individual Reliability	287	0.87	0.25	.58**	.40**	.47**	.53**	.51**	.50**	--		
8. Maximum Long String	287	6.68	4.89	.44**	.43**	.30**	.34**	.41**	.00	.10	--	
9. Average Long String	287	3.84	2.87	.43**	.43**	.30**	.34**	.47**	-.08	.09	.84**	--
10. Survey Time	279	24.28	21.26	.01	.01	.03	.01	.05	-.08	-.18**	.15**	.14*

Note. *$p < .05$, **$p < .01$

Table 4.
Means, standard deviations, internal consistency reliabilities, and correlations of substantive variables

	N	M	SD	1	2	3	4	5	6	7	8	9	10
1. Extraversion	287	4.43	1.07	(.93)									
2. Agreeableness	287	5.25	0.83	.29**	(.90)								
3. Conscientiousness	287	4.80	0.85	.23**	.30**	(.90)							
4. Neuroticism	287	3.80	1.08	-.31**	-.17**	-.22**	(.93)						
5. Openness	287	4.99	0.79	.31**	.38**	.27**	-.30**	(.89)					
6. PA	284	5.13	0.88	.54**	.29**	.37**	-.31**	.39**	(.88)				
7. NA	284	3.17	0.93	-.31**	-.18**	-.26**	.73**	-.16*	-.28**	(.87)			
8. Family Social Support	283	5.14	1.49	.18**	.18**	.28**	-.15*	.05	.31**	-.13*	(.93)		
9. Life Satisfaction	282	4.60	1.28	.32**	.15*	.31**	-.41**	.14*	.41**	-.41**	.43**	(.85)	
10. Physical Symptoms	284	2.01	0.61	-.09	-.04	-.15*	.47**	-.02	-.07	.49**	-.17**	-.30**	(.84)

Note. *$p < .05$, **$p < .01$

Table 5.
Differences Between Warning and Control Conditions

		Warning		Control			
Measure	*N*	*M*	*SD*	*M*	*SD*	*t*	*d*
ICRA	287	10.62	1.41	9.70	2.80	3.56**	0.42
Bogus Items	287	8.57	1.17	7.55	2.44	4.57**	0.53
Instructed Resp. Items	287	2.88	0.44	2.51	0.92	4.43**	0.51
Self-Reported Diligence	287	6.21	0.94	5.69	1.35	3.81**	0.45
Psychometric Antonyms	287	-0.73	0.30	-0.64	0.41	2.23*	0.25
Psychometric Synonyms	287	0.69	0.15	0.66	0.24	0.93	0.15
Individual Reliability	287	0.90	0.15	0.85	0.32	1.82	0.10
Maximum Long String	287	6.26	3.90	7.07	5.65	1.40	0.17
Average Long String	287	3.71	2.35	3.97	3.28	0.78	0.09
Survey Time	279	34.72	22.90	33.87	19.67	0.33	0.04

Note. $*p < .05$, $**p < .005$ *(corrected for family-wise error)*

Table 6.
Results of Logistic Regression Analyses

	OR	95% CI	Wald	df	*p*
Experimental Condition	0.95	0.57 - 1.58	0.04	1	0.84
ICRA	0.98	0.88 - 1.09	0.15	1	0.70
Bogus Items	1.03	0.90 - 1.17	0.13	1	0.72
Instructed Response Items	1.08	0.76 - 1.54	0.18	1	0.67
Self-Reported Diligence	0.98	0.79 - 1.21	0.04	1	0.84
Psychometric Antonyms	0.50	0.22 - 1.18	2.51	1	0.11
Psychometric Synonyms	0.92	0.26 - 3.23	0.02	1	0.90
Individual Reliability	1.67	0.48 - 5.75	0.66	1	0.42
Maximum Long String	0.99	0.93 - 1.04	0.23	1	0.63
Average Long String	0.91	0.77 - 1.08	1.18	1	0.28
Survey Time	0.99	0.98 - 1.01	1.34	1	0.25

Table 7.
Differences in Cronbach's Alphas between top 75% and bottom 25% of scorers on IER indices.

	Extrav	Agree	Consc	Neur	Open	PA	NA	Fam Supp	Life Sat	Phys Sym
ICRA	-.08**	.01	-.09*	-.10**	-.16**	-.02	.04	-.04*	-.06	.10**
Bogus Items	-.07**	-.09*	-.08*	-.17**	-.15**	-.02	.10**	-.01	.02	.04
Instructed Response Items	-.10**	-.03	-.10*	-.07**	-.29**	.02	.00	-.03	-.08*	.14**
Self-Reported Diligence	-.06**	-.03	-.03	-.18**	-.06	.01	.03	.01	-.01	.08*
Psychometric Antonyms	-.03	.00	-.16**	-.06*	-.06	.03	.05	-.03	.01	.07*
Psychometric Synonyms	-.09**	-.04	-.09*	-.04	-.12**	-.06*	-.04	-.23**	-.28**	.02
Individual Reliability	-.06**	-.06	-.12**	-.09**	-.12**	-.07*	-.15**	-.07**	-.27**	.03
Maximum Long String	.00	.01	.00	-.03	-.06*	.06*	.09**	.04*	.08*	.13**
Average Long String	-.02	.00	-.04	-.02	.00	.07*	.03	.05**	.10**	.05
Survey Time	.01	.05	-.06*	.00	.01	.09*	-.07*	.03	.05	.00

Note. *$p < .05$, **$p < .005$ *(corrected for family-wise error)*

Table 8.
Comparison of Cronbach's Alphas between top 75% and bottom 25% of scorers on IER indices.

	Substantive Scales with Reverse-Scored Items			Substantive Scales Scored in One Direction		
	Higher Among Bottom 25% of Scorers	Lower Among Bottom 25% of Scorers	No Significant Difference	Higher Among Bottom 25% of Scorers	Lower Among Bottom 25% of Scorers	No Significant Difference
ICRA	0%	60%	40%	20%	0%	80%
Bogus Items	0%	60%	40%	20%	0%	80%
Instructed Response Items	0%	60%	40%	20%	0%	80%
Self-Reported Diligence	0%	40%	60%	0%	0%	100%
Psychometric Antonyms	0%	20%	80%	0%	0%	100%
Psychometric Synonyms	0%	40%	60%	0%	40%	60%
Individual Reliability	0%	80%	20%	0%	60%	40%
Maximum Long String	0%	0%	100%	40%	0%	60%
Average Long String	0%	0%	100%	40%	0%	60%
Survey Time	0%	0%	100%	20%	0%	80%

Table 9.
Differences in Cronbach's alphas between experimental conditions.

Substantive Scale	Warning Condition	Control Condition	*W*	*p*
Extraversion	.93	.93	1.00	.501
Agreeableness	.91	.89	0.82	.117
Conscientiousness	.91	.88	0.75*	.044
Neuroticism	.93	.92	0.88	.214
Openness	.90	.88	0.83	.140
Positive Affectivity	.88	.88	1.00	.501
Negative Affectivity	.87	.86	0.93	.330
Family Social Support	.93	.93	1.00	.501
Life Satisfaction	.86	.85	0.93	.341
Physical Symptoms	.80	.87	0.65**	.005

Note. $*p < .05$, $**p < .005$ *(corrected for family-wise error)*

Table 10.

Summary of the results of regression analyses within the control condition across fifteen substantive predictor-criterion relationships moderated by each IER index.

	Total R^2	Avg. ΔR^2	Min ΔR^2	Max ΔR^2	SD ΔR^2	Avg. *p*	Min *p*	Max *p*	# Significant ΔR^2	# Significant ΔR^2 after correction
Extrav. x Family Social Support	.039	.016	.000	.034	.011	.254	.026*	.911	2	0
Extrav. x Life Satisfaction	.110	.007	.000	.028	.009	.471	.036*	.915	1	0
Extrav. x Physical Symptoms	.097	.003	.000	.008	.003	.585	.244	.894	0	0
Agree. x Family Social Support	.067	.008	.000	.043	.013	.583	.011*	.994	1	0
Consc. x Life Satisfaction	.090	.008	.001	.030	.010	.448	.031*	.783	1	0
Consc. x Physical Symptoms	.102	.012	.001	.036	.011	.378	.022*	.927	1	0
Neur. x Family Social Support	.035	.006	.000	.018	.005	.423	.107	.826	0	0
Neur. x Life Satisfaction	.235	.010	.000	.036	.013	.402	.010*	.897	2	0
Neur. x Physical Symptoms	.273	.009	.000	.024	.009	.346	.034*	.911	1	0
PA x Family Social Support	.099	.010	.001	.051	.015	.403	.005*	.704	1	0
PA x Life Satisfaction	.138	.019	.000	.055	.016	.195	.002**	.815	3	1
PA x Physical Symptoms	.090	.004	.000	.011	.004	.563	.170	.957	0	0
NA x Family Social Support	.051	.015	.000	.028	.009	.222	.041*	.942	1	0
NA x Life Satisfaction	.223	.016	.000	.058	.023	.433	.001**	1.000	2	2
NA x Physical Symptoms	.293	.017	.000	.041	.014	.234	.005*	.961	3	0

Note. $*p < .05$, $**p < .003$ *(corrected for family-wise error)*

Table 11.

Summary of the results of regression analyses within the control condition for fifteen substantive predictor-criterion relationships across IER index moderators.

	Total R^2	Avg. ΔR^2	Min ΔR^2	Max ΔR^2	SD ΔR^2	Avg. *p*	Min *p*	Max *p*	# Significant ΔR^2	# Significant ΔR^2 after correction
ICRA	.137	.011	.000	.043	.013	.385	.011*	.915	2	0
Bogus Items	.143	.007	.000	.024	.007	.393	.052	.821	0	0
Instructed Response	.170	.015	.001	.055	.015	.300	.002**	.919	3	1
Diligence	.110	.010	.000	.041	.013	.427	.005*	.911	3	0
Psychometric Antonyms	.114	.008	.000	.021	.008	.474	.043*	.957	1	0
Psychometric Synonyms	.146	.018	.001	.058	.016	.245	.001**	.762	5	1
Individual Reliability	.160	.012	.000	.057	.014	.318	.001**	.804	1	1
Max Long String	.102	.011	.000	.036	.011	.382	.022*	.994	2	0
Avg. Long String	.104	.007	.000	.040	.011	.470	.006*	.897	1	0
Survey Time	.107	.007	.000	.051	.013	.567	.005*	1.000	1	0

Note. *$p < .05$, **$p < .003$ *(corrected for family-wise error)*

Table 12.
Results of regression analyses for fifteen substantive predictor-criterion relationships moderated by condition.

	ΔR^2	*p*
Extraversion x Family Social Support	.011	.072
Extraversion x Life Satisfaction	.001	.594
Extraversion x Physical Symptoms	.001	.570
Agreeableness x Family Social Support	.003	.347
Conscientiousness x Life Satisfaction	.000	.760
Conscientiousness x Physical Symptoms	.001	.997
Neuroticism x Family Social Support	.001	.702
Neuroticism x Life Satisfaction	.007	.120
Neuroticism x Physical Symptoms	.002	.411
Positive Affectivity x Family Social Support	.001	.586
Positive Affectivity x Life Satisfaction	.005	.201
Positive Affectivity x Physical Symptoms	.003	.344
Negative Affectivity x Family Social Support	.001	.651
Negative Affectivity x Life Satisfaction	.002	.386
Negative Affectivity x Physical Symptoms	.002	.449

APPENDIX A

International Personality Item Pool – 100 Item Version.

Items
1. Am the life of the party.
2. Insult people.
3. Am always prepared.
4. Get stressed out easily.
5. Have a rich vocabulary.
6. Often feel uncomfortable around others.
7. Am interested in people.
8. Leave my belongings around.
9. Am relaxed most of the time.
10. Have difficulty understanding abstract ideas.
11. Feel comfortable around people.
12. Am not interested in other people's problems.
13. Pay attention to details.
14. Worry about things.
15. Have a vivid imagination.
16. Keep in the background.
17. Sympathize with others' feelings.
18. Make a mess of things.
19. Seldom feel blue.
20. Am not interested in abstract ideas.
21. Start conversations.
22. Feel little concern for others.
23. Get chores done right away.
24. Am easily disturbed.
25. Have excellent ideas.
26. Have little to say.
27. Have a soft heart.
28. Often forget to put things back in their proper place.
29. Am not easily bothered by things.
30. Do not have a good imagination.
31. Talk to a lot of different people at parties.
32. Am not really interested in others.
33. Like order.
34. Get upset easily.
35. Am quick to understand things.
36. Don't like to draw attention to myself.
37. Take time out for others.
38. Shirk my duties.
39. Rarely get irritated.
40. Try to avoid complex people.
41. Don't mind being the center of attention.

42. Am hard to get to know.
43. Follow a schedule.
44. Change my mood a lot.
45. Use difficult words.
46. Am quiet around strangers.
47. Feel others' emotions.
48. Neglect my duties.
49. Seldom get mad.
50. Have difficulty imagining things.
51. Make friends easily.
52. Am indifferent to the feelings of others.
53. Am exacting in my work.
54. Have frequent mood swings.
55. Spend time reflecting on things.
56. Find it difficult to approach others.
57. Make people feel at ease.
58. Waste my time.
59. Get irritated easily.
60. Avoid difficult reading material.
61. Take charge.
62. Inquire about others' well-being.
63. Do things according to a plan.
64. Often feel blue.
65. Am full of ideas.
66. Don't talk a lot.
67. Know how to comfort others.
68. Do things in a half-way manner.
69. Get angry easily.
70. Will not probe deeply into a subject.
71. Know how to captivate people.
72. Love children.
73. Continue until everything is perfect.
74. Panic easily.
75. Carry the conversation to a higher level.
76. Bottle up my feelings.
77. Am on good terms with nearly everyone.
78. Find it difficult to get down to work.
79. Feel threatened easily.
80. Catch on to things quickly.
81. Feel at ease with people.
82. Have a good word for everyone.
83. Make plans and stick to them.
84. Get overwhelmed by emotions.
85. Can handle a lot of information.
86. Am a very private person.
87. Show my gratitude.

88. Leave a mess in my room.
89. Take offense easily.
90. Am good at many things.
91. Wait for others to lead the way.
92. Think of others first.
93. Love order and regularity.
94. Get caught up in my problems.
95. Love to read challenging material.
96. Am skilled in handling social situations.
97. Love to help others.
98. Like to tidy up.
99. Grumble about things.
100. Love to think up new ways of doing things.

APPENDIX B

Positive Affectivity Negative Affectivity Scale (PANAS).

Items
1. Indicate to what extent you generally feel interested.
2. Indicate to what extent you generally feel distressed.
3. Indicate to what extent you generally feel excited.
4. Indicate to what extent you generally feel upset.
5. Indicate to what extent you generally feel strong.
6. Indicate to what extent you generally feel guilty.
7. Indicate to what extent you generally feel scared.
8. Indicate to what extent you generally feel hostile.
9. Indicate to what extent you generally feel enthusiastic.
10. Indicate to what extent you generally feel proud.
11. Indicate to what extent you generally feel irritable.
12. Indicate to what extent you generally feel alert.
13. Indicate to what extent you generally feel ashamed.
14. Indicate to what extent you generally feel inspired.
15. Indicate to what extent you generally feel nervous.
16. Indicate to what extent you generally feel determined.
17. Indicate to what extent you generally feel attentive.
18. Indicate to what extent you generally feel jittery.
19. Indicate to what extent you generally feel active.
20. Indicate to what extent you generally feel afraid.

APPENDIX C

Family Social Support Scale.

Items
1. Over the last semester, how much has your family acted in ways that show they appreciate what you do?
2. Over the last semester, how much has your family treated you with respect?
3. Over the last semester, how much has your family given you useful information and advice when you wanted it?
4. Over the last semester, how much has your family helped out when too many things needed to get done?
5. Over the last semester, how much has your family listened when you wanted to confide about things that were important to you?

APPENDIX D

Satisfaction with Life Scale.

Items
1. In most ways my life is close to my ideal.
2. The conditions of my life are excellent.
3. I am satisfied with my life.
4. So far I have gotten the important things I want in life.
5. If I could live my life over, I would change almost nothing.

APPENDIX E

Physical Symptoms Inventory – 12-item Version.

Items
1. How often have you experienced an upset stomach or nausea?
2. How often have you experienced trouble sleeping?
3. How often have you experienced headache?
4. How often have you experienced acid indigestion or heartburn?
5. How often have you experienced eye strain?
6. How often have you experienced diarrhea?
7. How often have you experienced stomach cramps (not menstrual)?
8. How often have you experienced constipation?
9. How often have you experienced ringing in the ears?
10. How often have you experienced loss of appetite?
11. How often have you experienced dizziness?
12. How often have you experienced tiredness or fatigue?

APPENDIX F

Combined Bogus Item Scale/IER Scale.

Items
1. I have never brushed my teeth.
2. I sleep less than one hour per night.
3. I have been to every country in the world.
4. I do not understand a word of English.
5. I have never used a computer.
6. I am using a computer currently.
7. I can run two miles in two minutes.
8. I am enrolled in a Psychology class currently.
9. I work 14 months in a year.

APPENDIX G

Diligence Scale.

Items
1. I carefully read every survey item.
2. I should've paid closer attention to the items than I did.
3. I probably should have been more careful during this survey.
4. I worked to the best of my abilities in this study.
5. I put forth my best effort in responding to this survey.
6. I didn't give this survey the time it deserved.
7. I was dishonest on some items.
8. I was actively involved in this study.
9. I rushed through this survey.

APPENDIX H

Instructed Response Items

Items
1. Please select neither agree nor disagree.
2. Please select strongly agree.
3. Please select strongly disagree.

APPENDIX I

Item Content Recall Approach (ICRA) Scale.

Target Items
1. I believe that I could have a satisfying career working as a librarian.
2. I enjoy listening to classical music.
3. I would like to go skydiving.
4. I like the taste of Brussels sprouts.
5. I would enjoy living in Alaska during the wintertime.
6. I think stamp collecting would be a fun hobby.
7. If my friends dared me to eat a live goldfish, I would probably do it.
8. I have had a recurring dream in which all my teeth have fallen out.
9. I have a fear of spiders.
10. I would be happy spending an afternoon at an art museum.
11. I would be impatient if I had to wait in line at an amusement park ride.
Quiz Items
1. Which of the following occupations were you asked about earlier in this questionnaire?
2. Which type of music were you asked about earlier in this questionnaire?
3. Which of the following "extreme" sports were you asked about earlier in this questionnaire?
4. Which of the following vegetables were you asked about earlier in this questionnaire?
5. Which U.S. State were you asked about earlier in this questionnaire?
6. Which of the following hobbies were you asked about earlier in this questionnaire?
7. Earlier in this questionnaire, we asked you about eating _____ as part of a dare.
8. Earlier in the questionnaire, we asked you if you had experienced a recurring dream about _____.
9. Earlier in the questionnaire, we asked you whether you had a fear of _____.
10. Earlier in the questionnaire, we asked you whether you would like to spend an afternoon at _____.
11. Earlier in the questionnaire, we asked you whether you would be impatient if you had to wait in line at _____.