Journal of Applied Psychology
2015, Vol. 100, No. 3, 828–845

0021-9010/15/$12.00 http://dx.doi.org/10.1037/a0038510

Insufficient Effort Responding: Examining an Insidious Confound in Survey Data

Jason L. Huang and Mengqiao Liu
Wayne State University

Nathan A. Bowling
Wright State University

Insufficient effort responding (IER; Huang, Curran, Keeney, Poposki, & DeShon, 2012) to surveys has largely been assumed to be a source of random measurement error that attenuates associations between substantive measures. The current article, however, illustrates how and when the presence of IER can produce a systematic bias that *inflates* observed correlations between substantive measures. Noting that inattentive responses as a whole generally congregate around the midpoint of a Likert scale, we propose that $M_{attentive}$, defined as the mean score of attentive respondents on a substantive measure, will be negatively related to IER's confounding effect on substantive measures (i.e., correlations between IER and a given substantive measure will become less positive [or more negative] as $M_{attentive}$ increases). Results from a personality questionnaire (Study 1) and a simulation (Study 2) consistently support the hypothesized confounding influence of IER. Using an employee sample (Study 3), we demonstrated how IER can confound bivariate relationships between substantive measures. Together, these studies indicate that IER can inflate the strength of observed relationships when scale means depart from the scale midpoints, resulting in an inflated Type I error rate. This challenges the traditional view that IER attenuates observed bivariate correlations. These findings highlight situations where IER may be a methodological nuisance, while underscoring the need for survey administrators and researchers to deter and detect IER in surveys. The current article serves as a wake-up call for researchers and practitioners to more closely examine IER in their data.

Keywords: insufficient effort responding, careless responding, random responding, data screening and cleaning, response effort

> It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so.
>
> —Mark Twain

Survey measures are commonly used in organizational research to assess a variety of variables, including perceptions of the work environment, job attitudes, and employee personality (Spector, 1994). The use of questionnaires assumes that respondents carefully follow instructions and remain attentive throughout the survey. Insufficient effort responding (IER), which subsumes both occasional careless (Schmitt & Stults, 1985) and intentional random responding (Buechley & Ball, 1952), refers to a survey response set in which a person responds to items without sufficient regard to the content of the items and/or survey instructions (Huang, Curran, Keeney, Poposki, & DeShon, 2012; also see Meade & Craig, 2012). Researchers have entertained the general notion of IER in various research contexts such as personality testing (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), job analysis (Green & Stutzman, 1986), needs assessment ratings (Calsyn & Winter, 1999), and online survey research (Behrend, Sharek, Meade, & Wiebe, 2011). Recent advances in IER research have identified methods that enable effective detection of IER (Huang, Bowling, Liu, & Li, 2014; Huang et al., 2012; Meade & Craig, 2012).

Much of the current understanding of IER's impact focuses on scales' measurement properties. Regarded as a source of random measurement error, IER can reduce scale reliability (Huang et al., 2012). Furthermore, the presence of IER can disturb the factor structure of measures that include both positively and negatively worded items. As shown in simulation studies (Schmitt & Stults, 1985; Woods, 2006), the inclusion of as little of 10% careless responses can introduce a nuisance factor consisting of negatively worded items. Taken as a whole, the view of IER as a source of random measurement error has perpetuated the assumption that IER necessarily *attenuates* observed relationships between substantive variables (McGrath, Mitchell, Kim, & Hough, 2010).

This article was published Online First December 15, 2014.

Jason L. Huang and Mengqiao Liu, Department of Psychology, Wayne State University; Nathan A. Bowling, Department of Psychology, Wright State University.

An earlier version of this article was presented at the annual conference of the Society for Industrial and Organizational Psychology in Honolulu, Hawaii in 2014. We thank Larry Williams, Boris Baltes, Bob Vandenberg, and Amy Fan for valuable suggestions at various stages of this research. We appreciate guidance from Deidra Schleicher and two anonymous reviewers on the structure of the article. We also thank Ben Biermeier-Hanson, Ariel Lelchook, and Caleb Bragg for sharing their data for the pilot studies and Terry A. Beehr, Connie P. Watson, Amy McGinnis, Kim Porter, Shannon Rodriguez, and Srikrishnan Sundaresan for their effort in data collection in Study 3.

Correspondence concerning this article should be addressed to Jason L. Huang, Department of Psychology, Wayne State University, 5057 Woodward Ave., Detroit, MI 48202. E-mail: jasonhuang@wayne.edu

As we describe below—and in contrast to what has generally been assumed—IER may in many instances *inflate* associations among substantive measures, thus increasing Type I error rates. Consider, for instance, a researcher who hypothesizes a positive relationship between core self-evaluations and job satisfaction (see Judge, Locke, Durham, & Kluger, 1998). Unbeknownst to the researcher, some respondents engage in IER and *happen to* report lower levels of core self-evaluations and lower levels of job satisfaction than the other (attentive) respondents (see the next section for a detailed discussion on the mechanism for IER's confounding effect). Because of the presence of IER as a confound, the researcher observes a relationship between core self-evaluations and job satisfaction that is stronger than the true underlying relationship. Consequentially, the researcher has greater than the nominal 5% probability to erroneously declare a significant association. This issue is important as IER is quite prevalent within self-report data—for instance, Meade and Craig (2012) found that 10% to 12% undergraduate students respond to their survey in a careless or haphazard manner. Given recent calls for methodological rigor in psychological and organizational research (e.g., Kepes & McDaniel, 2013; Simons, 2014), raising awareness to IER's confounding effects can help avert potential inconsistencies and controversies due to IER.

The goal of the current article is to describe the conditions under which the presence of IER in survey data can serve as a confound in substantive measures, thereby correcting the common misconception that IER necessarily attenuates observed relationships. Specifically, we argue that when IER is present, it often results in mean scores near the midpoint of a response scale for substantive measures. When attentive respondents score on average near either end of the response scale on the substantive measure, a phenomenon commonly seen in variables in organizational research (e.g., participants generally score above the scale midpoint on job satisfaction, but generally well below the scale midpoint on counterproductive work behavior), the observed scores on the substantive measure will be contaminated by the mean difference between IER and attentive individuals.

Our study of IER's potential confounding effect is important for two primary reasons. First, by addressing the misconception that IER serves to attenuate expected associations, the current investigation underscores the need to thoroughly examine IER in survey data. Indeed, our research makes a timely addition to two recent articles (Huang et al., 2012; Meade & Craig, 2012) that reviewed and evaluated various IER detection methods. Knowing the impact of IER *and* having the tools to deter and detect it (also see Huang et al., 2014), researchers and practitioners will be in a better position to obtain accurate results from their survey studies. Second, our research identifies IER as a source of method variance, therefore connecting IER research with a well-developed methodological literature (see Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Spector, 2006). Given the availability of IER indices, researchers may start modeling IER as a measurable method variable.

In what follows, we first illustrate the mechanism by which IER can introduce systematic variance in substantive measures. Using a personality questionnaire (Study 1) and a simulation (Study 2), we first test the hypothesized pattern of IER's confounding effects. We further demonstrate how IER may inflate bivariate relationships within an employee dataset (Study 3).

Hypothesized Effects of Insufficient Effort Responding on Substantive Variables

When IER occurs, be it the careless endorsement of a few items due to temporary inattentiveness or the egregious random selection of response options in a deliberate attempt to rush through the questionnaire, the response set introduces variance unrelated to a substantive measure's true scores. According to classical test theory, the presence of IER should reduce measurement reliability, as it adds a random source of variance in measurement. Indeed, Huang et al. (2012) reported that when respondents were instructed to engage in IER, scale reliability estimates decreased. Using normal survey instructions, Huang et al. also observed significant increases in reliability estimates after removing a small number of respondents suspected of engaging in IER.

The increase of *random* measurement error will attenuate the association between two measures (Spearman, 1904). Consistent with this notion, a qualitative review by McGrath et al. (2010) based on six studies concluded that indices assessing random or careless responding can moderate the association between two conceptually related measures, such that the association is stronger for attentive (i.e., non-IER) participants than for inattentive (i.e., IER) participants. Thus, the accumulated knowledge base in the psychological literature appears to indicate that the presence of IER within a given dataset will, if anything, generally lead to the failure to detect an otherwise significant effect—a Type II error. As we present below, however, the presence of IER may in many cases introduce *systematic* variance, artificially inflate the observed relationships between substantive variables, and increase Type I error rates.

As we present in Figure 1, whether IER introduces systematic variance in survey results depends on the means of attentive ($M_{\text{attentive}}$) and inattentive (M_{IER}) participants on a given substantive scale (see the mathematical proof in the Appendix). Figure 1a, 1b, and 1c present scores on hypothetical measures on a 7-point Likert scale when $M_{\text{attentive}}$ is lower than, higher than, and identical to M_{IER}, respectively. To simplify this illustration, the continuous IER variable is dichotomized into two groups (attentive responding = 0, IER = 1). Across these examples, attentive data are plotted on the left and IER data are plotted on the right. Disentangling attentive data from IER data enables one to intuitively discern the impact of IER.

In Figure 1a, the attentive group as a whole scores lower on Hypothetical Measure A than the IER group. Thus, we observe a positive correlation between the dichotomous IER variable and scores on Measure A. Pooling scores across attentive and IER cases will result in Measure A's scores being confounded with IER. If attentive scores represent true scores and if IER scores are truly random, pooling attentive and IER cases together adds three sources of variance together: (a) true score variance due to attentive responses; (b) random error variance due to IER; and (c) systematic covariance *positively* associated with IER.

In Figure 1b, the reverse of Figure 1a occurs, with the attentive group scoring higher on Hypothetical Measure B than the IER group. That is, a negative correlation exists between IER and scores on Measure B. Thus, pooling scores across IER and the

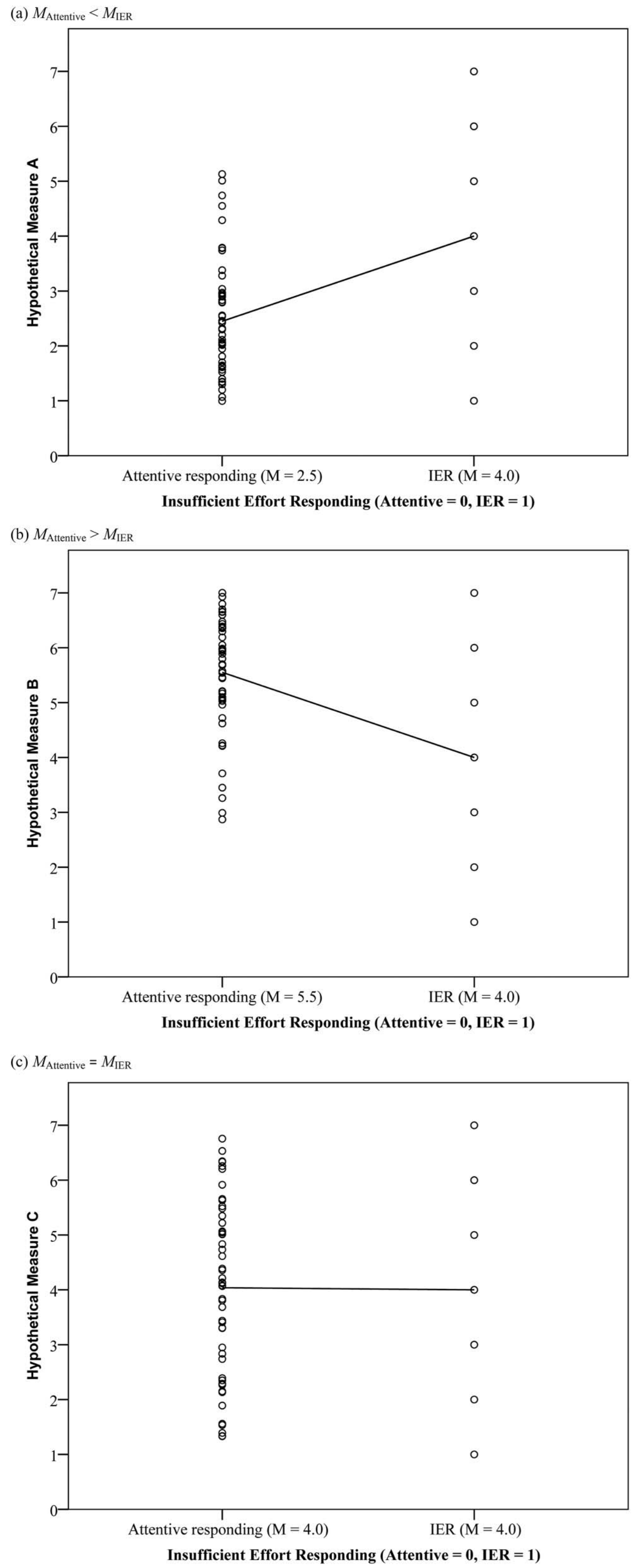

Figure 1. Illustration of process by which IER introduces nonzero variance in substantive measures. (a) $M_{\text{Attentive}} < M_{\text{IER}}$. (b) $M_{\text{Attentive}} > M_{\text{IER}}$. (c) $M_{\text{Attentive}} = M_{\text{IER}}$.

attentive groups will result in Measure B's scores being contaminated by IER in the direction opposite to Figure 1a. Assuming attentive scores are true scores and IER scores are random, pooling attentive and IER cases together again combines three sources of variance: (a) true score variance due to attentive responses; (b) random error variance due to IER; and (c) systematic covariance *negatively* associated with IER.

In Figure 1c, IER is unlikely to introduce systematic variance, because the attentive and IER groups have identical means on Hypothetical Measure C. That is, the presence of IER is not associated with scores on Measure C. In this case, to the extent that IER scores are random, consistent with classical test theory's assumption that true scores and error scores are uncorrelated, Measure C becomes less reliable owing to the presence of random measurement error.

A critical assumption for the current article—and a feature present in the examples in Figure 1—is that IER will result in scores that center around the substantive scale's midpoint. Three relevant IER scenarios support this assumption. First, when IER is truly random, a uniform distribution will result in a mean at the midpoint of the response scale. Second, when IER manifests in a particular pattern, such as long strings of responses (e.g., 15 *disagree*'s in a row; see Huang et al., 2012), the average responses will likely be at the midpoint across individuals who engage in IER. For example, across respondents engaging in IER, the probability of a long string of responses indicating agreement is identical to the probability of a long string of responses indicating disagreement. As a result, the average of long string responses across IER participants will tend to center around the midpoint of the response scale.

Third, within instances of careless responding, the average IER scores are pulled toward the midpoint *when* the average attentive responses are away from the midpoint. For example, on a typical 5-point scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*, if the mean of attentive responses is above the midpoint such as 4 (*agree*), there are more response categories for the occasional errors to occur below this mean (*strongly disagree*, *disagree*, and *neutral*) than above it (*strongly agree*). The converse is true if the mean of attentive responses is below the midpoint. If the mean of attentive responses is 2 (*disagree*), for example, then there are more response categories for errors to occur above than below this mean. The scenario is especially likely when there are reversed-worded items, which, when unheeded, will cancel out one's scores on the positively worded items. Taken together, absent other concurring response sets (e.g., acquiescence, impression management), various forms of IER are likely to result in scores on substantive measures averaging near the midpoint of the response scale.

With the assumption that IER will tend to produce scores near the scale midpoint on a given substantive measure, we illustrate the expected effects of IER with Figure 2. Five example values for $M_{\text{attentive}}$ are depicted, ranging from 2 to 6 on a 7-point scale. When $M_{\text{attentive}}$ is 2 and thus much lower than the midpoint, IER is positively correlated with scores on the substantive measure. Conversely, when $M_{\text{attentive}}$ is 6 and thus much higher than the midpoint, IER is negatively correlated with scores on the substantive measure. Taken together, the correlation between IER and the substantive measure is expected to: (a) become increasingly more negative as $M_{\text{attentive}}$ increases from the midpoint; (b) become

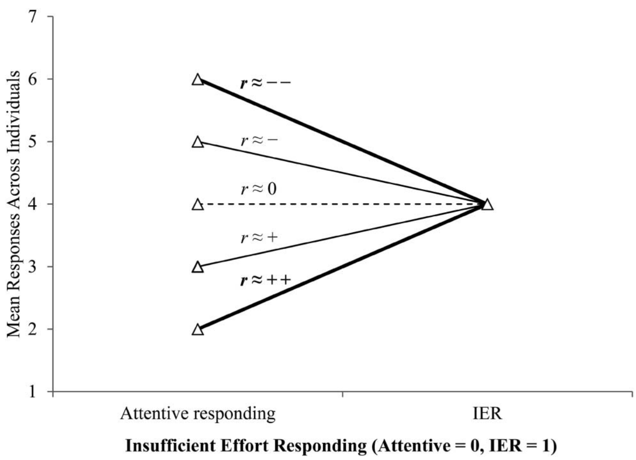

Figure 2. Hypothesized effect of IER based on mean of substantive measures. *Note*. *r* indicates the correlation between IER and scores on a substantive measure. **- -** < - < 0 < + < **+ +**.

increasingly more positive as $M_{\text{attentive}}$ decreases from the midpoint; and (c) be close to zero when $M_{\text{attentive}}$ is near the midpoint of the response scale. In other words, we expect a negative correlation between a substantive measure's $M_{\text{attentive}}$ and IER's confounding influence on the substantive measure.

The expectation stated above based on a 7-point scale can be applied to any Likert-type response scale. When a study includes substantive measures assessed on different response scales, $M_{\text{attentive}}$ may not be comparable across measures. A mean of 3.5, for example, is below the midpoint on a 7-point scale, and yet an identical mean of 3.5 is above the midpoint on a 5-point scale. When this occurs, different response scales can be rescaled by converting the lowest response category to 0 and the highest response category to 1, such that the *rescaled means* are comparable across measures, with a midpoint of 0.5. With this additional consideration, we formally express the expected effect in a hypothesis:

> *Hypothesis 1:* The correlation between IER and a substantive measure will be negatively correlated with the substantive measure's $M_{\text{attentive}}$ (rescaled when different response scales are used).

Three additional considerations are in order regarding the hypothesized influence of IER. First, it would be a moot point if the survey measures in the literature generally had means at the midpoints of their respective scales; in such situations, IER will simply contribute to measurement error. To provide an assessment of the literature, we conducted a cursory examination of 20 randomly selected articles published in *Journal of Applied Psychology* in 2012, focusing on the absolute difference between a measure's mean and its scale's midpoint. Of the 102 measures assessed on a 5-point scale, the average absolute difference was 0.68 (SD = 0.38). For measures assessed on a 6-point and 7-point scales, the average absolute difference was .94 (SD = .57, N = 9) and .99 (SD = .53, N = 25), respectively. Therefore, it appears rather common that measures used in applied research often yield means that differ from the midpoints of their response scales. Indeed, both core self-evaluations and job satisfaction, which we refer to in our earlier example, tend to have means higher than the midpoint of a response scale (e.g., Bowling & Hammond, 2008; Bowling, Hoepf, LaHuis, & Lepisto, 2013; Greenbaum, Mawritz, & Eissa, 2012; Judge, Erez, Bono, & Thoresen, 2003).

Second, when the presence of IER introduces systematic bias to substantive measures that have $M_{\text{attentive}}$ away from the midpoints of their response scales, the existence of IER as a common source of variance will then confound the association between these measures. In Figure 3, we illustrate how IER introduces common variance between two uncorrelated hypothetical measures. In Figure 3a, absent any IER, two substantive measures X and Y are

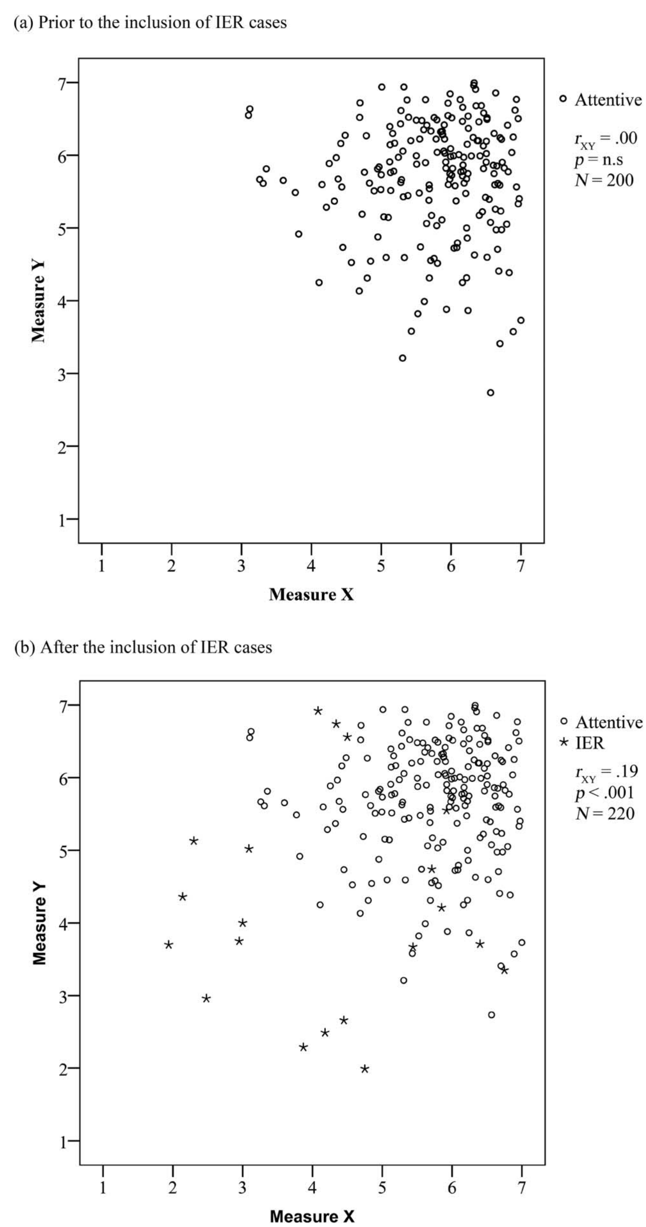

Figure 3. Illustration of process by which IER inflates association between substantive measures. (a) Before the inclusion of IER cases. (b) After the inclusion of IER cases. *Note*. Among IER cases, r_{XY} = .00, *ns*, N = 20.

uncorrelated, $r = .00$, *ns*, with means above the midpoint of the 7-point Likert scale. With the presence of IER in Figure 3b; however, the negative correlation between IER and each measure served as a common source of variance that leads to an inflated correlation between X and Y. As a result, a significant relationship, $r = .19$, $p < .001$ is observed between X and Y simply because both variables are contaminated with IER variance.

Third, it may not always be possible to accurately classify attentive and IER cases in survey data, making it difficult to assess $M_{\text{attentive}}$. However, research indicates that the large majority of respondents—perhaps 85% to 90%—engage in attentive responding (Huang et al., 2012; Meade & Craig, 2012), thus suggesting that the mean across all participants is a reasonable proxy for $M_{\text{attentive}}$.

Given these considerations, it is then possible to examine IER's confounding effect in survey data with substantive measures' means away from the midpoints of their response scales. For substantive measures such as core self-evaluations and job satisfaction that have means above the scales' midpoints, we can expect IER to be negatively related to these measures. In contrast, IER is expected to be positively related to substantive measures with means below the scales' midpoints, such as Machiavellianism and counterproductive work behavior (for information about the means of these scales, see Fox, Spector, Goh, Bruursema, & Kessler, 2012; Linton & Power, 2013; Spector et al., 2006; Zagenczyk et al., 2013). Thus, IER can serve as a confound for the observed bivariate associations. In the examples above, the positive relationships observed between core self-evaluations and job satisfaction and between Machiavallianism and CWB will both be artificially inflated as the result of IER. Accordingly, controlling for IER can reduce IER's confounding effect between substantive measures.

> *Hypothesis 2:* When substantive measures have means away from scale midpoints, partialling out IER will significantly change the magnitude of correlations between these measures.

We conducted three studies to examine Hypotheses 1 and 2 above. First, we tested Hypothesis 1 using responses to a personality questionnaire (Study 1) and a simulation (Study 2). We further examined Hypothesis 2 in an employee survey (Study 3).

In Study 1, we used Huang et al.'s (2012, Study 2) dataset to test for Hypothesis 1. Specifically, the study included 30 personality variables administered under normal instructions as well as several operationalizations of IER. We saw three benefits of using this published dataset. First, the study enabled us to examine the critical assumption that IER participants will score near the midpoints of response scales on substantive measures. Using cutoff scores empirically developed in an experiment, Huang et al. (2012) demonstrated significant improvement in internal consistency and unidimensionality in personality measures after removing 5% of total cases suspected of IER. If our assumption is supported, the 5% IER participants should score closer to the midpoints than the 95% presumed attentive respondents.

Second, the study allowed for the test of Hypothesis 1 with a reasonable sample size ($N = 30$ substantive measures). It would be challenging to have a large sample size from a survey study to test for the hypothesis, as each substantive measure represents one data point for the analysis. Although a sample size of 30 appears small, results from two pilot studies (*r*s = −.44 and −.96; details available from the first author) suggested that $N = 30$ would yield a power between 70% and 100% for an α rate of .05. Finally, we followed Meade and Craig's (2012) recommendation to assess IER with multiple detection indices. In so doing, we could establish the relationships between the IER response set and substantive measures while ruling out the potential concern that the relationships might be index-specific.

Study 1

Method

Participants. As described in Huang et al. (2012), the sample consisted of 345 students (68% female, on average 25 years old) enrolled in undergraduate psychology courses. Students participated in the online survey in exchange for extra course credit.

Measures

Personality scales. Measures of 30 Five Factor Model personality facets (10 items per facet; average α = .79) were presented on 12 Web pages, with each Web page containing 25 personality items. Responses were made on a 5-point scale (1 = *very inaccurate*; 5 = *very accurate*).

Insufficient effort responding. Huang et al. (2012, Study 2) operationalized IER with five indices: psychometric antonym, individual reliability, long string index, page time, and postsurvey self-report (see Table 1). They further found a single factor underlying all continuous IER indices (excluding long string index, which was dichotomous). To best capture the underlying IER construct and to avoid potential idiosyncratic effects attributable to any single IER index, we operationalized IER in Study 1 using the first factor scores saved from exploratory factor analysis on the continuous IER indices, with a higher score indicating greater likelihood of IER behavior.

Results

Table 2 presents descriptive statistics for each of the 30 personality facets, as well as correlation between the IER factor and each facet. The sizable correlations between the IER factor and each facet (i.e., $r_{\text{IER-X}}$) are worth noting, ranging from −.64 to .32, with 19 negative correlations and 7 positive ones reaching statistical significance. As an initial step to testing Hypothesis 1, which required the use of $M_{\text{attentive}}$, we proceeded to create two groups, with 5% of cases (see Huang et al., 2012) scoring highest on the IER factor identified as the IER group and the rest of the sample classified into the attentive group. The means for each group, as well as independent samples *t* tests for the difference between the two groups (see Table 2) lent credence to the assumption that IER would in general result in scores that center around the scale midpoint of a substantive measure. Specifically, the IER group scored significantly closer to the midpoint of "3" than the attentive group on 25 out of the 30 personality facets, with the exception of the five facets where attentive respondents also scored close to the midpoint of the response scale. We further tested this observed probability (i.e., 25 out of 30) against a binomial distribution with

Table 1
Description of IER Indices in Study 1

IER index	Detection approach	Steps for index construction
Psychometric antonym	Inconsistency	(a) Empirically identifying 30 item pairs sharing strongest negative correlations; (b) Computing within-person correlation across such items ($N = 30$).
Individual reliability	Inconsistency	(a) Separating items on each scale into odd- versus even-numbered ones; (b) Calculating half-scale scores for odd- and even-numbered items separately; (c) Computing within-person correlation across corresponding half-scale scores; (d) Applying Spearman-Brown split-half formula.
Long string	Response pattern	(a) Identifying highest number of consecutive endorsement of each response option; (b) Compare each number against an empirically developed cutoff.
Page time	Response time	(a) Recording the time between the initiation and submission of each substantive survey page; (b) Log transforming the average time across pages.
Postsurvey self-report	Self-report	(a) Including three quality assurance items (e.g., "I responded carelessly to the questions") after the main survey; (b) Averaging responses across items.

Note. See Huang et al. (2012) for details.

50% proportion and found significant support for the current assumption, $p < .001$.

Hypothesis 1 states that the associations between IER and substantive measures will be negatively correlated with substantive measures' $M_{attentive}$. To prepare for the analysis, we performed Fisher r to z transformation on the 30 $r_{\text{IER-X}}$ correlation coefficients.[1] A correlation between $M_{\text{attentive}}$ values and the z-transformed correlations supported Hypothesis 1, $r = -.96$, $p < .001$, $N = 30$. That is, the higher the attentive respondents scored on a personality facet, the more negative (or less positive) the facet's correlation with IER tended to be.[2]

We suggested above that the mean across all participants may serve as a reasonable proxy for $M_{\text{attentive}}$. Indeed, the correlation between M and $M_{\text{attentive}}$ was approaching unity, $r = .9999$, strongly supporting this notion. Thus, when only a small proportion of respondents engage in IER, one may use M instead of $M_{\text{attentive}}$ to assess IER's potential influence.

Discussion

Results for Study 1 provided strong support for Hypothesis 1 that the correlations between IER and substantive measures will be negatively related to attentive respondents' average scores on the substantive measures. When $M_{\text{attentive}}$ was above the scale midpoint (e.g., achievement facet of conscientiousness), IER was negatively associated with the substantive measure (e.g., $r_{\text{IER-achievement}} = -.41$). When $M_{\text{attentive}}$ was below the scale midpoint (e.g., vulnerability facet of neuroticism), IER was positively associated with the substantive measure (e.g., $r_{\text{IER-vulnerability}} = .27$). The pattern of IER's relationships with substantive measures in Study 1 suggests that IER may confound the observed relationships between substantive measures.

The use of a nonexperimental cross-sectional research design in Study 1 limits our ability to draw causal inferences, as one might surmise that some unobserved variables could have caused the particular pattern of results we found. An unsuspecting researcher, for instance, might interpret the associations between IER and personality measures in Study 1 substantively, concluding that participants engaging in IER tended to be less conscientious, agreeable, extraverted, open, and more neurotic. To establish the causal inference that IER can introduce systematic variance in substantive measures in the condition we outlined above, we used simulated data in Study 2. Specifically, we aimed to demonstrate how the presence of 10% of *completely random responses* in a dataset may introduce spurious correlations *among otherwise unrelated variables*. The percentage of random responses (10% of total cases) we adopted is consistent with previous simulation studies (Schmitt & Stults, 1985; Woods, 2006) as well as with estimates of IER prevalence found among actual participants (Meade & Craig, 2012). We also provide estimates based on 5% of all cases as random responses (a more conservative scenario) for comparison purposes without hypothesis testing.

Study 2

Method

Data generation. Before generating the data, we determined that a sample size of 100,000 would be sufficient to produce stable results without discernible variation resulting from sampling error. We adopted a 7-point scale (ranging from 1 to 7) as the response scale for the simulation, and manipulated the means for attentive cases to range from 2.0 to 6.0 with 0.1 unit increment, resulting in 41 simulated substantive variables. A power analysis based on estimates from the two pilot studies and from Study 1 indicated that a sample size of 41 would have power ranging from 85% to 100%.

For the 90,000 attentive cases, we generated data on the 41 variables using random normal distribution with predetermined means and a SD of 1. When the variables' means were close to either end of the response scale, some of the data from random

[1] In the present article, we z-transformed all $r_{\text{IER-X}}$ before all tests of Hypothesis 1. We thank an anonymous reviewer for bringing this issue to our attention.

[2] Similar support for Hypothesis 1 was found when IER was operationalized with each of the continuous IER indices, rs ranging from $-.91$ to $-.95$, $p < .001$. Furthermore, Hypothesis 1 was supported with two additional operationalizations of IER: (a) when IER was measured with psychometric synonym index (see Meade & Craig, 2012), $r = -.96$, $p < .001$; and (b) when IER was classified using the long string index (see Huang et al., 2012), $r = -.81$, $p < .001$. Detailed results are available from the first author.

Table 2
IER's Associations With Substantive Measures in Study 1

Measure	M	SD	$r_{\text{IER-X}}$	$M_{\text{Attentive}}$ (N = 327)	M_{IER} (N = 18)	t
A-altruism	4.01	0.56	−.58***	4.06	3.05	13.23***
A-cooperation	3.38	0.61	−.19***	3.40	3.07	3.45**
A-modesty	2.94	0.60	.17***	2.93	3.11	−2.55*
A-morality	3.64	0.61	−.25***	3.66	3.21	3.11**
A-sympathy	3.45	0.55	−.23***	3.47	3.08	2.98**
A-trust	3.46	0.69	−.36***	3.49	2.91	5.81***
C-achievement	3.71	0.64	−.41***	3.75	2.98	5.11***
C-cautiousness	3.05	0.66	−.01	3.05	3.15	−1.39
C-dutifulness	3.85	0.53	−.54***	3.90	2.98	14.12***
C-orderliness	3.34	0.73	−.13*	3.36	2.98	4.37***
C-self-discipline	2.91	0.76	−.05	2.90	2.98	−1.01
C-self-efficacy	3.73	0.56	−.56***	3.78	2.94	10.06***
E-activity level	3.01	0.44	−.05	3.01	2.99	0.28
E-assertiveness	3.58	0.69	−.36***	3.62	2.94	7.00***
E-cheerfulness	3.94	0.60	−.64***	4.00	2.71	10.08***
E-excitement seeking	3.56	0.66	−.32***	3.60	2.90	4.46***
E-friendliness	3.83	0.66	−.47***	3.88	2.92	10.05***
E-gregariousness	3.69	0.67	−.36***	3.72	3.08	8.88***
N-anger	2.80	0.75	.20***	2.79	3.01	−2.32*
N-anxiety	2.99	0.72	.14*	2.99	3.02	−0.33
N-depression	2.40	0.76	.32***	2.37	2.83	−4.39***
N-immoderation	3.24	0.66	.04	3.25	3.08	2.01
N-self-consciousness	2.67	0.67	.25***	2.65	2.92	−3.90***
N-vulnerability	2.66	0.68	.27***	2.64	2.96	−3.10**
O-adventurousness	3.22	0.50	−.24***	3.23	3.02	4.14***
O-artistic interests	3.92	0.65	−.45***	3.97	3.02	15.44***
O-emotionality	3.66	0.59	−.31***	3.70	3.04	9.07***
O-imagination	3.68	0.61	−.41***	3.73	2.84	11.79***
O-intellect	3.57	0.64	−.33***	3.60	3.02	7.62***
O-liberalism	2.92	0.57	.18***	2.91	3.27	−2.70**

Note. N = 345. A = agreeableness; C = conscientiousness; E = extraversion; N = neuroticism; O = openness to experience; $r_{\text{IER-X}}$ = correlation between IER factor score and a substantive measure.
* $p < .05$. ** $p < .01$. *** $p < .001$.

normal distribution were out of the predefined bounds (i.e., <1 or >7). As survey respondents with extremely high or low scores on a construct would be bounded by the Likert scale provided, we attempted to simulate that response process, taking into account potential measurement error. Specifically, we replaced values that were smaller than 1 with values from a random normal distribution with a mean of 1.5 (SD = 0.15), and replaced values that were larger than 7 with values from another random normal distribution (M = 6.5, SD = 0.15). The resulting means of the 41 variables ranged from 2.16 to 5.84 in the attentive sample (see Table 3), indicating successful manipulation of the measures' means.

For the 10,000 IER cases, we simulated random responses on each of the 41 variables using uniform distribution ranging from 1 to 7. In the IER subsample, the means of the random responses on the 41 variables ranged from 3.96 to 4.03.

Results

As data for the attentive and IER subsamples were generated randomly, scores on the 41 measures should be unrelated within each subsample. We first examined the relationships among measures within each subsample to ensure the validity of data manipulation. In the attentive subsample (N = 90,000), the correlation between any two variables ranged from −.01 to .01, while in the IER subsample (N = 10,000), the corresponding range of correlation was −.03 to .04. The average association across all measures within each subsample was .00.

We further used exploratory factor analysis (EFA) with principal axis factoring to examine whether there was any pattern in the correlations, focusing on two related statistics: initial communality and first factor loading. Initial communality in an EFA indicates the amount of variance in a focal variable accounted for by all other variables, whereas first factor loading indicates the correlation between a focal variable and a latent factor that maximally explains nonunique variance in all variables. As both subsamples consisted of random data, we expected low values on both statistics within each subsample. Results confirmed our expectation (Table 3): The maximum and average initial communality were .001 and .000 for the attentive subsample and .006 and .004 for the IER subsample, whereas the first factor loadings ranged from −.08 to .06 and −.09 to .14, respectively, providing evidence for the validity of data manipulation.

To examine the impact of IER, we proceeded to pool the attentive and IER subsamples together (N = 100,000). Correlation analysis revealed higher associations among measures in the pooled data than in the attentive subsample, with the correlation

Table 3
IER's Confounding Effects in Simulated Data in Study 2

	Attentive subsample (N = 90,000)				IER subsample (N = 10,000)				Pooled total sample (N = 100,000)				
Variable	M	SD	h^2	λ	M	SD	h^2	λ	M	SD	h^2	λ	$r_{\text{IER-X}}$
M2.0	2.16	0.78	.001	.02	4.00	1.74	.004	−.04	2.35	1.08	.22	.52	.51
M2.1	2.23	0.81	.000	.04	3.99	1.74	.003	−.03	2.41	1.08	.20	.49	.49
M2.2	2.31	0.84	.000	−.01	3.98	1.74	.003	.00	2.48	1.09	.17	.46	.46
M2.3	2.39	0.86	.000	.06	4.01	1.73	.005	−.03	2.55	1.09	.16	.44	.44
M2.4	2.48	0.88	.000	.00	4.01	1.74	.004	−.04	2.63	1.10	.15	.42	.42
M2.5	2.56	0.90	.001	.01	4.01	1.72	.004	.00	2.71	1.10	.13	.39	.40
M2.6	2.65	0.91	.000	−.04	4.01	1.73	.005	.07	2.79	1.10	.11	.37	.37
M2.7	2.74	0.92	.001	.06	4.02	1.73	.006	.13	2.87	1.10	.10	.35	.35
M2.8	2.83	0.94	.000	.04	4.00	1.74	.004	.04	2.95	1.10	.08	.32	.32
M2.9	2.92	0.96	.000	−.01	4.02	1.74	.003	.01	3.03	1.11	.07	.30	.30
M3.0	3.02	0.96	.001	−.03	4.01	1.73	.003	−.03	3.12	1.10	.06	.27	.27
M3.1	3.12	0.97	.000	.01	4.03	1.74	.003	.05	3.21	1.10	.05	.25	.25
M3.2	3.21	0.97	.000	.01	3.99	1.73	.004	.04	3.29	1.10	.04	.21	.21
M3.3	3.30	0.98	.000	.00	3.98	1.75	.002	−.01	3.37	1.10	.03	.19	.19
M3.4	3.41	0.98	.000	.04	3.99	1.73	.005	−.07	3.47	1.10	.02	.16	.16
M3.5	3.51	0.99	.000	.02	4.01	1.74	.005	.04	3.56	1.09	.02	.14	.14
M3.6	3.61	0.99	.001	.03	3.99	1.73	.005	−.07	3.65	1.09	.01	.11	.11
M3.7	3.69	0.99	.000	.00	3.98	1.72	.004	.06	3.72	1.09	.01	.08	.08
M3.8	3.80	0.99	.000	.03	3.97	1.73	.004	.05	3.82	1.09	.00	.05	.05
M3.9	3.90	0.99	.000	.02	3.99	1.73	.004	.00	3.91	1.09	.00	.02	.02
M4.0	4.00	0.99	.000	.00	3.97	1.72	.006	.14	4.00	1.09	.00	−.01	−.01
M4.1	4.09	1.00	.000	.01	4.00	1.74	.004	−.04	4.08	1.09	.00	−.03	−.02
M4.2	4.19	0.99	.001	−.01	4.00	1.73	.004	−.07	4.17	1.09	.00	−.06	−.05
M4.3	4.30	0.99	.000	.01	4.00	1.75	.005	.05	4.27	1.09	.01	−.08	−.08
M4.4	4.40	0.99	.001	−.06	4.01	1.73	.005	.00	4.36	1.09	.01	−.11	−.11
M4.5	4.50	0.99	.001	−.08	4.02	1.74	.005	−.06	4.45	1.09	.01	−.13	−.13
M4.6	4.60	0.98	.001	.04	4.02	1.74	.005	.05	4.54	1.09	.02	−.16	−.16
M4.7	4.70	0.98	.000	.04	3.99	1.73	.005	.06	4.63	1.10	.03	−.19	−.19
M4.8	4.79	0.97	.000	−.01	4.01	1.74	.003	−.03	4.71	1.10	.04	−.21	−.21
M4.9	4.89	0.97	.000	.02	4.01	1.72	.005	.02	4.80	1.10	.05	−.24	−.24
M5.0	4.98	0.96	.000	.04	3.99	1.74	.004	−.01	4.88	1.11	.06	−.27	−.27
M5.1	5.08	0.95	.000	.03	4.03	1.73	.004	−.07	4.98	1.10	.07	−.28	−.29
M5.2	5.17	0.94	.001	−.02	3.99	1.74	.004	.02	5.05	1.11	.08	−.32	−.32
M5.3	5.26	0.92	.000	.03	4.01	1.74	.005	.11	5.13	1.10	.09	−.34	−.34
M5.4	5.35	0.91	.001	−.05	4.00	1.74	.005	−.08	5.21	1.10	.11	−.37	−.37
M5.5	5.44	0.90	.000	−.01	3.96	1.73	.003	−.02	5.29	1.11	.13	−.40	−.40
M5.6	5.52	0.88	.000	−.04	3.99	1.74	.005	−.09	5.37	1.10	.14	−.42	−.42
M5.7	5.61	0.85	.000	.01	4.00	1.73	.004	.02	5.45	1.09	.16	−.44	−.44
M5.8	5.69	0.83	.000	−.01	4.01	1.74	.004	−.01	5.52	1.09	.18	−.47	−.46
M5.9	5.77	0.81	.000	.00	4.02	1.71	.005	.10	5.59	1.08	.19	−.49	−.49
M6.0	5.84	0.78	.000	.00	4.02	1.74	.004	.01	5.66	1.07	.21	−.51	−.51

Note. Naming of variables indicates the mean of the initial random normal distribution used to generate attentive responses. For example, M3.8 indicates the 90,000 attentive cases were originally generated from a random normal distribution with a mean of 3.8. h^2 = initial communality in EFA, which is the R^2 from multiple regression analysis predicting the focal variable from all the other variables; λ = loading on the first factor extracted from EFA, which is the correlation between a variable and the latent factor that maximally explains nonunique variance among all observed variables; $r_{\text{IER-X}}$ = correlation between the dichotomous IER indicator (0 = attentive responding, 1 = IER) and observed scores on a variable.

between any two measures ranging from −.26 to .26 (see Table 4, below the diagonal, for correlations among selective variables). An EFA further revealed the confounding effect of IER: Maximum and average initial communality increased to .22 and .08, respectively, whereas first factor loadings ranged from −.51 to .52 (Table 3).

To test Hypothesis 1, we first created a dichotomous variable labeled IER (attentive responding = 0 and IER = 1) and correlated this variable with each hypothetical measure. As the correlation ($r_{\text{IER-X}}$) in Table 3 indicated, sizable effects emerged. After Fisher r to z transformation, the correlation $r_{\text{IER-X}}$ was perfectly associated with $M_{\text{attentive}}$, $r = -1.00$, $p < .001$, $N = 41$. Further, $r_{\text{IER-X}}$ was nearly identical to the measures' loadings on the first factor from EFA ($r = 1.00$, $p < .001$, $N = 41$), indicating that IER was the confounding factor among observed measures. Most important, controlling for the dichotomous IER indicator, the patterns of association among substantive measures were completely removed, with first-order partial correlations among measures ranging from −.01 to .01 (N = 100,000).

Finally, for comparison purposes, we created a different pooled sample consisting of all 90,000 attentive cases and 4,737 IER cases (5% of the entire sample of 94,737 cases). As shown above the diagonal in Table 4, even with only 5% of the data being IER, the confounding effect was still present. Where the

Table 4
Spurious Correlations Introduced by 10% and 5% of IER Cases in Data in Study 2

	M2.0	M2.2	M2.4	M2.6	M2.8	M3.0	. . .	M5.0	M5.2	M5.4	M5.6	M5.8	M6.0
M2.0	—	.16	.15	.12	.11	.09	. . .	−.08	−.11	−.12	−.14	−.17	−.18
M2.2	.24	—	.12	.11	.09	.07	. . .	−.07	−.09	−.10	−.12	−.14	−.16
M2.4	.22	.19	—	.10	.08	.07	. . .	−.07	−.08	−.10	−.11	−.13	−.14
M2.6	.19	.17	.16	—	.07	.06	. . .	−.06	−.07	−.08	−.10	−.11	−.12
M2.8	.17	.15	.13	.11	—	.05	. . .	−.05	−.07	−.07	−.08	−.09	−.11
M3.0	.14	.12	.11	.10	.09	—	. . .	−.04	−.05	−.06	−.07	−.08	−.09
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
M5.0	−.13	−.12	−.11	−.10	−.09	−.07	. . .	—	.05	.06	.07	.08	.09
M5.2	−.16	−.14	−.13	−.12	−.11	−.08	. . .	.08	—	.07	.08	.10	.10
M5.4	−.19	−.17	−.16	−.14	−.12	−.10	. . .	.10	.11	—	.10	.11	.12
M5.6	−.21	−.19	−.18	−.16	−.13	−.11	. . .	.11	.13	.16	—	.13	.14
M5.8	−.24	−.21	−.20	−.17	−.15	−.13	. . .	.13	.14	.17	.20	—	.16
M6.0	−.26	−.24	−.22	−.18	−.16	−.13	. . .	.14	.16	.19	.21	.23	—

Note. For correlations below the diagonal, N = 100,000, consisting of 10% of IER and 90% of attentive responses. For correlations above the diagonal, N = 94,737, consisting of 5% of IER and 95% of attentive responses. Correlations among M3.1 to M4.9 were omitted. The full correlation table is available upon request from the first author.

true correlation should be zero, the observed correlations ranged from −.18 to .16, as compared with −.26 to .24 when 10% of the cases as IER. This pattern of spurious correlations highlights the importance of examining and detecting IER in survey measures even when most of participants are believed to respond attentively.

Discussion

Study 2 replicated the finding from Study 1 using simulated data. Most critically, Study 2 demonstrated that the presence of 10% (and even 5%) IER cases can *cause* spurious relationships among otherwise uncorrelated measures. In so doing, we established the mechanism by which IER confounds associations among substantive measures in survey data.

Our simulation finding that IER can inflate the correlation between two substantive measures extends beyond past simulations, which either focused on IER's negative impact on scale measurement properties (Schmitt & Stults, 1985; Woods, 2006) or evaluated approaches to detecting IER (Meade & Craig, 2012; Pinsoneault, 2002). Rather than reinforcing the assumption that IER represents random "noise" that generally attenuates observed relationships, findings from Study 2 indicate otherwise: under the conditions specified above, researchers may find *stronger* associations between substantive measures when they fail to control for IER. Ignoring the effects of IER, therefore, may in some instances cause one to make a Type I error.

Given the findings from Studies 1 and 2, we progress to examining IER's potential influence in a field survey in Study 3. Thus, we can directly examine Hypothesis 2 about changes in observed correlations after partialling out IER. Using an archival dataset previously published in Bowling, Beehr, Bennett, and Watson (2010), Study 3 included substantive variables that are commonly measured with self-report in organizational research, such as job satisfaction, interpersonal treatment, and trait affectivity. More important, we anticipated attentive respondents to have varying rescaled means on these variables (e.g., high rescaled means for job satisfaction and low rescaled means for negative affectivity), making it possible to test for IER's confounding effects.

Study 3

Method

Participants. Participants were recruited from nonfaculty employees of a university in the Midwestern United States. Each participant was given the choice between completing an electronic or hardcopy version of the study questionnaire. Nearly all participants completed the electronic version of the questionnaire. Three steps were taken to encourage participation: (a) cosponsorship of the project was obtained from the University's Human Resources Department, which publicized the study questionnaire and actively encouraged participation; (b) prospective participants received a reminder e-mail approximately 2 weeks after the initial participation invitation; and (c) as an incentive to participate, those who responded to the study questionnaire were entered into a raffle to win dinner at a local restaurant.

The sample of 466 respondents was predominantly White (91%) and female (76%), with an average age of 43 (SD = 10) and average tenure of 14 years with the university (SD = 2.5). Twenty-six percent of the participants held supervisory positions.

Measures

Job satisfaction. Job satisfaction (α = .91) was assessed with five items from Brayfield and Rothe (1951). A sample item is "I find real enjoyment in my work." Responses were made on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Life satisfaction. Life satisfaction (α = .89) was assessed with five items from Diener, Emmons, Larsen, and Griffin (1985). A sample item is, "I am satisfied with my life." Each item was rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Coworker treatment. Coworker treatment (α = .80) was measured with four items from Donovan, Drasgow, and Munson

(1998). A sample item is "Coworkers treat me with respect." Ratings were made on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Supervisor treatment. The supervisor treatment measure (α = .90) consisted of four items written to parallel the Donovan et al.'s (1998) coworker treatment items. A sample item is "My supervisor treats me with respect." Each item was rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Work centrality. Work centrality (α = .85) was measured with 12 items from Paullay, Alliger, and Stone-Romero (1994). A sample item is "Work should be considered central to life." Responses were made on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

Core self-evaluations. Core self-evaluations (α = .82) was assessed with 12 items from Judge et al. (2003). A sample item is "I am confident I get the success I deserve in life." Each item was rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Positive and negative affectivity. Positive affectivity (PA) and negative affectivity (NA; αs = .91 and .86) were assessed with 10 items each from Watson, Clark, and Tellegen (1988). Each item asked participants to describe how they generally feel. Sample items include "Excited" (PA) and "Upset" (NA). The items were rated on a 5-point scale from 1 (*very slightly or not at all*) to 5 (*extremely*).

Insufficient effort responding. We assessed IER with an individual reliability index (Jackson, 1976, as cited in Johnson, 2005), which outperformed other IER measures in Meade and Craig's (2012) simulation study (they referred to it as the "Even-Odd Consistency" index). Specifically, we calculated half-measure scores using the odd-numbered and even-numbered items from each measure. Individual reliability was based on the within-person correlation between the odd and even half-measures of the eight substantive measures (larger values on the individual reliability index indicate higher within-person consistency and presumably lower levels of IER; see Huang et al., 2012). So that higher scores on the IER index would indicate a greater degree of IER, we rescored the individual reliability index by multiplying it by −1 (see Huang et al., 2014).

Results

Table 5 presents descriptive statistics and intercorrelations for all variables in Study 3. These correlations among substantive measures are consistent with past research findings. Job satisfaction and life satisfaction, for instance, were positively related to each other; coworker treatment, supervisor treatment, core self-evaluations, and PA were positively related and NA was negatively related to job satisfaction (for similar findings, see Judge et al., 1998; Tait, Padgett, & Baldwin, 1989; Thoresen, Kaplan, Barsky, Warren, & de Chermont, 2003; Viswesvaran, Sanchez, & Fisher, 1999). More important, the reversed individual reliability index was significantly associated with all of the substantive measures except work centrality. The significant correlations ranged from .18 to .49 in absolute values.

We should note that individual reliability in the current study was very high, with only 0.9% of respondents' scores falling below the .30 cutoff suggested by Jackson (1976; cf. Huang et al., 2012). This may suggest that the respondents as a whole were quite attentive, and that relatively higher scores on the reversed individual reliability index may indicate unintentional carelessness rather than deliberate speeding through the survey. Consistent with the argument presented above, we used the observed means of substantive measures (i.e., *M*) as proxies for $M_{\text{attentive}}$, and examined the pattern of associations between $r_{\text{IER-X}}$ and rescaled *M*. The directions of the significant correlations conform to the pattern depicted in Figure 2. Consistent with Hypothesis 1, we found a significant negative correlation between a substantive measure's rescaled *M* and IER's association with the substantive measure, $r = -.96$, $p < .001$, $N = 8$.

Hypothesis 2 states that when substantive measures have means away from scale midpoints, partialling out IER will change the magnitude of correlations between these measures. To quantify the change in effect size from zero-order correlation to first-order partial correlation, we computed the magnitude of the change and applied Olkin and Finn's (1995, Model C, p. 160) formula to

Table 5
Descriptive Statistics and Correlations for Study 3 Variables

	1	2	3	4	5	6	7	8	9
1. Job satisfaction	**.91**								
2. Life satisfaction	.34***	**.89**							
3. Coworker treatment	.33***	.27***	**.80**						
4. Supervisor treatment	.54***	.21***	.48***	**.90**					
5. Work centrality	.28***	.01	−.04	.07	**.85**				
6. Core self-evaluations	.46***	.59***	.31***	.38***	.11*	**.82**			
7. Positive affectivity	.43***	.48***	.21***	.28***	.16***	.58***	**.91**		
8. Negative affectivity	−.35***	−.40***	−.31***	−.31***	−.10*	−.61***	−.32***	**.86**	
9. Individual reliability reversed	−.41***	−.29***	−.44***	−.42***	.00	−.48***	−.18***	.49***	—
M	5.31	4.83	5.67	5.65	3.22	5.27	3.73	1.56	−0.89
SD	1.21	1.32	1.01	1.35	0.92	0.77	0.65	0.50	0.13
Likert scale categories	7	7	7	7	7	7	5	5	—
Rescaled *M*	0.72	0.64	0.78	0.78	0.37	0.71	0.68	0.14	—

Note. N = 466. Rescaled M = (M – 1)/(Categories – 1).
* $p < .05$. ** $p < .01$. *** $p < .001$. Cronbach's α presented on the diagonal in bold.

assess if such a change was significant. A significant change would suggest that IER introduced nonzero systematic variance among observed variables and thus support Hypothesis 2.

As shown in Table 6, 20 out of the 28 correlations among substantive measures (71%) displayed significant changes in magnitude after controlling for IER. Removing IER as a confound resulted in an average reduction of effect size of .08 (SD = .06). Thus, Hypothesis 2 received general support. Indeed, nonsignificant changes from zero-order correlations to first-order partial correlations were mostly observed for work centrality, which had no association with IER (perhaps because it had a $M_{attentive}$ near the scale midpoint).

Discussion

As the individual reliability index indicated overall low levels of IER, Study 3 served as a relatively conservative test of our hypotheses. Even with the high levels of attentiveness, we found sizable associations between IER and substantive measures. For example, IER had a negative correlation with core self-evaluations, $r = -.48$, $p < .001$ and job satisfaction, $r = -.41$, $p < .001$. Supporting Hypothesis 2, partialling out IER significantly changed the observed relationships among substantive variables. Following the same examples, removing variance due to IER reduced the correlation between core self-evaluations and job satisfaction from .46 to .33. Clearly, if unheeded, IER may serve as a confound and thus produce systematically biased research conclusions.

General Discussion

Across three studies, we demonstrated nonzero correlations between IER and various substantive measures, and we found that the magnitude and direction of such correlations were associated with attentive respondents' average score on a substantive measure. Following Lykken's (1968) constructive replication approach, we found consistent support for the hypothesized confounding effect across diverse research contexts and survey measures using various operationalizations of IER. Furthermore, Study 3 indicates that partialling out scores on an IER index may remove the confounding effects of IER. Taken together, the current findings help delineate when and how the presence of IER can confound survey results. That is, IER is likely to inflate—rather than attenuate—observed correlations between substantive measures when observed scale means depart from the scale midpoints.

The current article makes a timely addition to emerging research on IER. Specifically, a series of recent articles (Huang et al., 2012, 2014; Meade & Craig, 2012) have evaluated various methods to detect IER. However, before the current article, the consequence of including IER in one's data remained poorly understood, with the misconception that IER invariably acts as a source of random measurement error in survey data. This misconception may lead unsuspecting researchers to be complacent about the need to screen for IER when they find support for their hypotheses. That is, upon finding significant results, researchers who mistakenly assume that IER necessarily results in attenuated effects—and thus Type II errors—would incorrectly conclude that they obtained statistically significant results *despite* the potential presence of IER. As we have shown in the current article, however, they may have in fact obtained statistically significant results *because of* the effects of IER. By showing IER can have differential impact on survey data depending on properties of the substantive measures, the current article helps make a much stronger case for researchers to perform diligent checks for IER in their data.

The need for increased appreciation of the potential effects of IER is shown in a recent survey of SIOP members (Liu, Bowling, Huang, & Kent, 2013), which revealed that less than half of the respondents routinely check for IER in their survey data. Indeed, very few published research articles on substantive areas in organizational research reported screening for IER (for exceptions, see Chiaburu, Huang, Hutchins, & Gardner, 2014; Seo & Barrett, 2007). The biasing effect of IER identified in the current article across survey contexts suggests that perhaps the topic of IER has received insufficient attention from researchers and practitioners.

In light of the current findings, we believe that assessing the presence of IER in survey data *before* hypothesis testing should be a standard practice. We especially caution against screening for IER after failing to find support for hypotheses, as this can increase Type I error rate. Instead, researchers should decide how they will address IER (e.g., what IER indices to use, what specific IER cut-off scores to set, etc.) *before* conducting their primary data analysis (see Huang, 2014, for an example). In fact, safeguarding against IER should occur before and during data collection. Some IER detection approaches such as survey time and infrequency scales (Huang et al., 2012, 2014; Meade & Craig, 2012) require

Table 6
Partial Correlations Among Study 3 Variables Controlling for IER

	1	2	3	4	5	6	7	8
1. Job satisfaction	—	.09***	.15***	.10***	−.03	.13***	.03*	−.16***
2. Life satisfaction	.25***	—	.10***	.11***	.00	.05**	.03**	−.09***
3. Coworker treatment	.19***	.17***	—	.12***	.00	.18***	.06**	−.19***
4. Supervisor treatment	.44***	.11*	.36***	—	−.01	.15***	.05**	−.18***
5. Work centrality	.31***	.01	−.05	.08	—	−.02	.00	.02
6. Core self-evaluations	.33***	.54***	.13**	.23***	.12**	—	.01	−.12***
7. Positive affectivity	.40***	.45***	.15**	.23***	.17***	.57***	—	−.05*
8. Negative affectivity	−.19***	−.30***	−.12*	−.13**	−.12*	−.49***	−.27***	—

Note. N = 466. Partial correlations controlling for the reversed individual reliability index are presented below the diagonal. Changes from zero-order correlation to first-order partial correlation are presented above the diagonal, calculated as ($r_{xy} - r_{xy.IER}$), with significance levels estimated based on Model C from Olkin and Finn (1995).
* $p < .05$. ** $p < .01$. *** $p < .001$.

planning before survey administration. Furthermore, the survey administrator should strive to properly motivate respondents to provide quality data (e.g., by issuing a benign warning about data screening; Huang et al., 2014).

Implications for Future Research

By challenging the prevailing assumption that IER is always a source of random measurement error, the current research provides compelling evidence that IER can inflate the strength of observed relationships by introducing common method variance. In that regard, our finding echoes Williams and Brown (1994) in that the use of a common method (i.e., self-report survey) can introduce sizable correlations when there is indeed no underlying relationship. However, unlike the other sources of common method variance (see Podsakoff et al., 2003 for a summary), the mechanisms through which IER produces systematic error variance depends on the means of attentive respondents on the substantive measures. Thus, our finding confirmed Spector's (2006) observation that reliance on self-report does not automatically inflate observed correlations because of the common method. Instead, the presence of a small percentage of IER (e.g., as little as 5% of the total sample), combined with an $M_{\text{attentive}}$ near either end of the response scale, can inflate observed correlations and should be considered as a special case of method bias. With the availability of multiple IER detection approaches (Huang et al., 2012, 2014; Meade & Craig, 2012), survey researchers may now model IER as a specific unique method factor (see Williams, Hartman, & Cavazotte, 2010).

The current finding that NA was contaminated by IER ($r = .49$ in Study 3) points to an interesting direction for common method variance research, as NA has often been examined as a biasing factor (Spector, 2006). If agreement to NA items is in part tainted by high IER behavior, controlling for scores on NA may inadvertently control for IER. Thus, the current findings suggest that previous studies regarding the biasing effects of NA should be re-examined to first rule out potential influence of IER.

A notable limitation of the current article is the focus on IER's influence in self-report survey data measured at a single time point. In that regard, our article simply offers additional evidence to the commonly shared notion that empirical findings based on cross-sectional self-report data warrant great scrutiny (see Kozlowski, 2009). Nevertheless, by raising attention to IER, our article lays the groundwork for a better understanding of IER's influence in other types of research designs. We highlight three areas for future research. First, researchers may examine IER in conjunction with nonself-report variables. For example, when employees report personality variables and supervisors rate their job performance, one may expect IER in personality measures to generally act as measurement error and attenuate expected associations between personality traits and job performance (see McGrath et al., 2010). Second, researchers may start to investigate the stability of IER behavior across surveys. That is, a few respondents may engage in IER behavior repeatedly in different waves of a longitudinal study. How such consistent IER behavior may affect survey results needs further examination. Finally, the spurious relationships between IER and substantive variables prevented us from identifying individual characteristics that give rise to IER behavior. Future research can use observer-reported personality (e.g., Oh, Wang, & Mount, 2011) as predictors of IER in self-report surveys.

The current research points to the need to better understand survey contexts. Specifically, we call for researchers to move beyond the type of samples (e.g., undergraduates, employees; Liu et al., 2013) and to consider contextual features that may influence IER behavior. One such contextual feature is the perceived obligation to complete the survey. Undergraduate students may participate in surveys to earn extra credit or fulfill course obligation (Meade & Craig, 2012), whereas employees may perceive pressure from supervisors to complete a workplace questionnaire. It is conceivable that under such circumstances, IER may occur because respondents are motivated to complete the survey without clear incentive for data quality.

A second contextual feature worth exploring is survey payment. Paid participants may range from community or organizational samples responding to surveys in exchange for monetary rewards, to crowdsourcing survey respondents such as Amazon's MTurk (Mechanical Turk). Of interest to the authors, mixed results were reported regarding data quality from MTurk, with some studies supporting its use (e.g., Buhrmester, Kwang, & Gosling, 2011), whereas others suggesting higher rates of IER within MTurk samples (e.g., Goodman, Cryder, & Cheema, 2013). One might conjecture that IER will occur less frequently when survey payment is somewhat contingent on data quality.

Another area in need of future research is the intersection between IER and other response sets, such as socially desirable responding and acquiescence (Huang et al., 2014). For instance, in organizational surveys where employees doubt the anonymity of survey results, IER behavior may result in acquiescence, that is, agreeing to positive attributes without reflecting on one's true perception. The extent to which IER coupled with other response sets will result in the same pattern of biasing effects remains to be investigated.

Practical Implications

The current article highlights the importance of screening for IER, even in situations where survey respondents are believed to be generally motivated and attentive. For organizational surveys, such as test validation, needs assessment, and multisource feedback, the presence of IER can exert unexpected effects for the unprepared survey administrator. The consequences of having IER in applied data can pose costly issues for organizations, such as falsely adopting an invalid selection test (or not adopting a valid test). Given the prevalence of self-reports in organizational settings, practitioners should be cognizant of the issues surrounding IER in survey administration/analysis.

In practice, one can use different operationalizations of IER depending on the context (see Huang et al., 2012, 2014; Meade & Craig, 2012). With an effective IER measure, applied survey administrators can first examine if IER is related to substantive measures with *M*s away from the midpoints of response scales. When evidence of biasing effect is found, practitioners can either control for the IER scores in subsequent analyses, or develop cutoffs rationally or empirically (Huang et al., 2012) to screen out extreme forms of IER.

Conclusion

We demonstrated that IER can serve as a source of common variance that confounds relationships between substantive measures. Contrary to the traditional view that IER necessarily attenuates observed bivariate relationships (e.g., McGrath et al., 2010), the current article showed that the presence of IER can inflate the magnitude of observed correlations between substantive measures when scale means depart from the scale midpoints. By showing that the systematic effects of IER are contingent upon the mean of attentive respondents on a substantive measure, we illuminated an additional mechanism through which IER can pose a threat on survey data quality.

References

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43,* 800–813. http://dx.doi.org/10.3758/s13428-011-0081-0

Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods, 7,* 400–417. http://dx.doi.org/10.1177/1094428104268542

Bowling, N. A., Beehr, T. A., Bennett, M. M., & Watson, C. P. (2010). Target personality and workplace victimization: A prospective analysis. *Work & Stress, 24,* 140–158. http://dx.doi.org/10.1080/02678373.2010.489635

Bowling, N. A., & Hammond, G. D. (2008). A meta-analytic examination of the construct validity of the Michigan Organizational Assessment Questionnaire Job Satisfaction Subscale. *Journal of Vocational Behavior, 73,* 63–77. http://dx.doi.org/10.1016/j.jvb.2008.01.004

Bowling, N. A., Hoepf, M. R., LaHuis, D. M., & Lepisto, L. R. (2013). Mean job satisfaction levels over time: Are things bad and getting worse? *The Industrial-Organizational psychologist, 50,* 57–64.

Brayfield, A. H., & Rothe, H. F. (1951). An index of job satisfaction. *Journal of Applied Psychology, 35,* 307–311. http://dx.doi.org/10.1037/h0055617

Buechley, R., & Ball, H. (1952). A new test of validity for the group MMPI. *Journal of Consulting Psychology, 16,* 299–301. http://dx.doi.org/10.1037/h0053897

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Calsyn, R. J., & Winter, J. P. (1999). Understanding and controlling response bias in needs assessment studies. *Evaluation Review, 23,* 399–417. http://dx.doi.org/10.1177/0193841X9902300403

Chiaburu, D. S., Huang, J. L., Hutchins, H. M., & Gardner, R. G. (2014). Trainees' perceived knowledge gain unrelated to the training domain: The joint action of impression management and motives. *International Journal of Training and Development, 18,* 37–52. http://dx.doi.org/10.1111/ijtd.12021

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71–75. http://dx.doi.org/10.1207/s15327752jpa4901_13

Donovan, M. A., Drasgow, F., & Munson, L. J. (1998). The Perceptions of Fair Interpersonal Treatment Scale: Development and validation of a measure of interpersonal treatment in the workplace. *Journal of Applied Psychology, 83,* 683–692. http://dx.doi.org/10.1037/0021-9010.83.5.683

Fox, S., Spector, P. E., Goh, A., Bruursema, K., & Kessler, S. R. (2012). The deviant citizen: Measuring potential positive relations between counterproductive work behaviour and organizational citizenship behaviour. *Journal of Occupational and Organizational Psychology, 85,* 199–220. http://dx.doi.org/10.1111/j.2044-8325.2011.02032.x

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26,* 213–224. http://dx.doi.org/10.1002/bdm.1753

Green, S. B., & Stutzman, T. (1986). An evaluation of methods to select respondents to structured job-analysis questionnaires. *Personnel Psychology, 39,* 543–564. http://dx.doi.org/10.1111/j.1744-6570.1986.tb00952.x

Greenbaum, R. L., Mawritz, M. B., & Eissa, G. (2012). Bottom-line mentality as an antecedent of social undermining and the moderating roles of core self-evaluations and conscientiousness. *Journal of Applied Psychology, 97,* 343–359. http://dx.doi.org/10.1037/a0025217

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581–595. http://dx.doi.org/10.1037/0021-9010.75.5.581

Huang, J. L. (2014). Does cleanliness influence moral judgments? Response effort moderates the effect of cleanliness priming on moral judgment. *Frontiers in Psychology, 5,* 1–8. http://dx.doi.org/10.3389/fpsyg.2014.01276

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2014). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. [Advance online publication]. *Journal of Business and Psychology*. Advance online publication. http://dx.doi.org/10.1007/s10869-014-9357-6

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

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 Web-based personality inventories. *Journal of Research in Personality, 39,* 103–129. http://dx.doi.org/10.1016/j.jrp.2004.09.009

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: Development of a measure. *Personnel Psychology, 56,* 303–331. http://dx.doi.org/10.1111/j.1744-6570.2003.tb00152.x

Judge, T. A., Locke, E. A., Durham, C. C., & Kluger, A. N. (1998). Dispositional effects on job and life satisfaction: The role of core evaluations. *Journal of Applied Psychology, 83,* 17–34. http://dx.doi.org/10.1037/0021-9010.83.1.17

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99,* 422–431. http://dx.doi.org/10.1037/0033-2909.99.3.422

Kepes, S., & McDaniel, M. A. (2013). How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 6,* 252–268. http://dx.doi.org/10.1111/iops.12045

Kozlowski, S. W. J. (2009). Editorial. *Journal of Applied Psychology, 94,* 1–4. http://dx.doi.org/10.1037/a0014990

Linton, D. K., & Power, J. L. (2013). The personality traits of workplace bullies are often shared by their victims: Is there a dark side to victims? *Personality and Individual Differences, 54,* 738–743. http://dx.doi.org/10.1016/j.paid.2012.11.026

Liu, M., Bowling, N. A., Huang, J. L., & Kent, T. (2013). Insufficient effort responding to surveys as a threat to validity: The perceptions and practices of SIOP members. *The Industrial-Organizational psychologist, 51,* 32–38.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151–159. http://dx.doi.org/10.1037/h0026141

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136,* 450–470. http://dx.doi.org/10.1037/a0019216

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

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96,* 762–773. http://dx.doi.org/10.1037/a0021832

Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin, 118,* 155–164. http://dx.doi.org/10.1037/0033-2909.118.1.155

Paullay, I. M., Alliger, G. M., & Stone-Romero, E. F. (1994). Construct validation of two instruments designed to measure job involvement and work centrality. *Journal of Applied Psychology, 79,* 224–228. http://dx.doi.org/10.1037/0021-9010.79.2.224

Pinsoneault, T. B. (2002). A variable response inconsistency scale and a true response inconsistency scale for the Millon Adolescent Clinical Inventory. *Psychological Assessment, 14,* 320–330. http://dx.doi.org/10.1037/1040-3590.14.3.320

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879–903. http://dx.doi.org/10.1037/0021-9010.88.5.879

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

Seo, M. G., & Barrett, L. F. (2007). Being emotional during decision making, good or bad? An empirical investigation. *Academy of Management Journal, 50,* 923–940. http://dx.doi.org/10.5465/AMJ.2007.26279217

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9,* 76–80. http://dx.doi.org/10.1177/1745691613514755

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15,* 72–101. http://dx.doi.org/10.2307/1412159

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior, 15,* 385–392. http://dx.doi.org/10.1002/job.4030150503

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9,* 221–232. http://dx.doi.org/10.1177/1094428105284955

Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviors created equal? *Journal of Vocational Behavior, 68,* 446–460. http://dx.doi.org/10.1016/j.jvb.2005.10.005

Tait, M., Padgett, M. Y., & Baldwin, T. T. (1989). Job and life satisfaction: A reevaluation of the strength of the relationship and gender effects as a function of the date of the study. *Journal of Applied Psychology, 74,* 502–507. http://dx.doi.org/10.1037/0021-9010.74.3.502

Thoresen, C. J., Kaplan, S. A., Barsky, A. P., Warren, C. R., & de Chermont, K. (2003). The affective underpinnings of job perceptions and attitudes: A meta-analytic review and integration. *Psychological Bulletin, 129,* 914–945. http://dx.doi.org/10.1037/0033-2909.129.6.914

Viswesvaran, C., Sanchez, J. I., & Fisher, J. (1999). The role of social support in the process of work stress: A meta-analysis. *Journal of Vocational Behavior, 54,* 314–334. http://dx.doi.org/10.1006/jvbe.1998.1661

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,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes, 57,* 185–209. http://dx.doi.org/10.1006/obhd.1994.1011

Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods, 13,* 477–514. http://dx.doi.org/10.1177/1094428110366036

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28,* 186–191. http://dx.doi.org/10.1007/s10862-005-9004-7

Zagenczyk, T., Cruz, K., Woodard, A., Walker, J. J., Few, W. W., Kiazad, K., & Raja, M. (2013). The moderating effect of Machiavellianism on the psychological contract breach-organizational identification/disidentification relationships. *Journal of Business and Psychology, 28,* 287–299. http://dx.doi.org/10.1007/s10869-012-9278-1

(*Appendix follows*)

Appendix

Proof that the Confounding Effect of IER Depends on Mean Difference between IER and Attentive Groups

In Table A, we start with two distinct original populations: the attentive population O′ with known sample size, mean and variance of N_0, $\mu_{O'}$, and $\sigma^2_{O'}$ and the IER population I′ with known sample size, mean and variance of N_1, $\mu_{I'}$, and $\sigma^2_{I'}$, respectively.

We consider the pooled population for observed score X a linear combination of two interim variables O and I: X = O + I. Both populations O and I have N observations. Population O consists of N_0 attentive scores and N_1 scores of 0, while population I consists of N_1 attentive scores and N_0 scores of 0. We further define p as proportion of IER cases in the pooled population ($p = N_1/N$) and q as proportion of attentive cases in the pooled population ($q = N_0/N$).

Following the equation for point biserial correlation:

$$r_{pb} = \frac{\mu_1 - \mu_0}{\sigma}\sqrt{\frac{N_1 N_0}{N^2}},$$

the correlation between dichotomous variable G (IER Group: attentive responding = 0, IER = 1) and continuous variable X (observed score) is:

$$r_{GX} = \frac{\mu_{I'} - \mu_{O'}}{\sigma_X}\sqrt{\frac{N_1 N_0}{N^2}}. \tag{1}$$

According to the formula above, it appears that the mean difference between the IER group and the attentive group (i.e., $\mu_{I'} - \mu_{O'}$) will influence the correlation between IER condition and observed scores, such that the correlation increases when the discrepancy between $\mu_{I'}$ and $\mu_{O'}$ becomes greater (everything else held constant). In other words, the observed scores are likely confounded by the presence of IER. However, because it is unclear how σ_X, the denominator, is affected by the mean difference between the IER group and the attentive group, additional derivation for σ_X is needed below.

$$\text{As X} = \text{O} + \text{I}, \sigma^2_X = \sigma^2_O + \sigma^2_I - 2Cov_{OI}. \tag{2}$$

To find out σ_X, we need to solve for each of the three components in σ^2_X above.

Table A
Pooling Attentive and IER Data

	Original populations		Interim populations		Pooled population	
Cases	Attentive	IER	Attentive	IER	Observed	IER condition
1	O_1	—	O_1	0	O_1	0
2	O_2	—	O_2	0	O_2	0
3	O_3	—	O_3	0	O_3	0
. . .	. . .	—	. . .	0	. . .	0
N_0	O_{N0}	—	O_{N0}	0	O_{N0}	0
1	—	I_1	0	I_1	I_1	1
2	—	I_2	0	I_2	I_2	1
3	—	I_3	0	I_3	I_3	1
. . .	—	. . .	0	. . .	. . .	1
N_1	—	I_{N1}	0	I_{N1}	I_{N1}	1
Name	O′	I′	O	I	X	G
Mean	$\mu_{O'}$	$\mu_{I'}$	μ_O	μ_I	μ_X	—
Variance	$\sigma^2_{O'}$	$\sigma^2_{I'}$	σ^2_O	σ^2_I	σ^2_X	—

(*Appendix continues*)

Following the computational formula for variance:

$$\sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N},$$

variance of the interim variable σ_O is:

$$\sigma_O^2 = \frac{\sum O^2 - \frac{(\sum O)^2}{N}}{N}. \tag{3}$$

As

$$\sigma_{O'}^2 = \frac{\sum O^2 - \frac{(\sum O)^2}{N_0}}{N_0} = \frac{\sum O^2 - \frac{(\sum O)^2}{qN}}{qN}$$

and

$$\mu_{O'} = \frac{\sum O}{N_0} = \frac{\sum O}{qN}$$

(where $q = N_0/N$), we can rearrange the equations and express ΣO^2 and ΣO in the following two equations:

$$\sum O^2 = qN \cdot \sigma_{O'}^2 + \frac{(\sum O)^2}{qN}, \tag{4}$$

and

$$\sum O = qN \cdot \mu_{O'}. \tag{5}$$

Thus, we can rewrite Equation 3 by substituting ΣO^2 and ΣO from Equations 4 and 5:

$$\begin{aligned} \sigma_O^2 &= \frac{qN \cdot \sigma_{O'}^2 + \frac{(\sum O)^2}{qN} - \frac{(\sum O)^2}{N}}{N} = \frac{qN \cdot \sigma_{O'}^2 + \frac{(qN \cdot \mu_{O'})^2}{qN} - \frac{(qN \cdot \mu_{O'})^2}{N}}{N} \\ &= q \cdot \sigma_{O'}^2 + q \cdot \mu_{O'}^2 - q^2 \cdot \mu_{O'}^2 \end{aligned} \tag{6}$$

Similarly,

$$\sigma_I^2 = p \cdot \sigma_{I'}^2 + p \cdot \mu_{I'}^2 - p^2 \cdot \mu_{I'}^2 \tag{7}$$

In addition, following the computational formula for covariance

$$Cov_{XY} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{N},$$

(*Appendix continues*)

we have

$$Cov_{OI} = \frac{\sum OI - \frac{\sum O \sum I}{N}}{N} = \frac{0 - \frac{\sum O \sum I}{N}}{N} = -\frac{\sum O \sum I}{N^2} \text{ (where} \sum OI = 0). \tag{8}$$

From Equation 5, we know:

$$\sum O = qN \cdot \mu_{O'}, \text{ and} \tag{9}$$

$$\sum I = pN \cdot \mu_{I'}, \tag{10}$$

Thus, Equation 8 can be rewritten by substituting ΣO and ΣI from Equations 9 and 10:

$$Cov_{OI} = -\frac{\sum O \sum I}{N^2} = -\frac{qN \cdot \mu_{O'} \cdot pN \cdot \mu_{I'}}{N^2} = pq \cdot \mu_{O'}\mu_{I'} \tag{11}$$

Given Equations 6, 7, and 11, we can re-express Equation 2, $\sigma_X^2 = \sigma_O^2 + \sigma_I^2 - 2Cov_{OI}$, as follows:

$$\begin{aligned} \sigma_X^2 &= q \cdot \sigma_{O'}^2 + q \cdot \mu_{O'}^2 - q^2 \cdot \mu_{O'}^2 + p \cdot \sigma_{I'}^2 + p \cdot \mu_{I'}^2 - p^2 \cdot \mu_{I'}^2 - 2pq \cdot \mu_{O'}\mu_{I'} \\ &= q \cdot \sigma_{O'}^2 + p \cdot \sigma_{I'}^2 + q(1-q) \cdot \mu_{O'}^2 + p(1-p) \cdot \mu_{I'}^2 - 2pq \cdot \mu_{O'}\mu_{I'} \\ &= q \cdot \sigma_{O'}^2 + p \cdot \sigma_{I'}^2 + pq \cdot \mu_{O'}^2 + pq \cdot \mu_{I'}^2 - 2pq \cdot \mu_{O'}\mu_{I'} \\ &= q \cdot \sigma_{O'}^2 + p \cdot \sigma_{I'}^2 + pq(\mu_{I'} - \mu_{O'})^2 \end{aligned} \tag{12}$$

We can re-express the point biserial correlation by substituting σ_X from Equation 12 as follows:

$$\begin{aligned} r_{GX} &= \frac{\mu_{I'} - \mu_{O'}}{\sigma_X} \sqrt{\frac{N_1 N_0}{N^2}} = \frac{\mu_{I'} - \mu_{O'}}{\sqrt{q \cdot \sigma_{O'}^2 + p \cdot \sigma_{I'}^2 + pq(\mu_{I'} - \mu_{O'})^2}} \sqrt{\frac{pNqN}{N^2}} \\ &= \sqrt{pq} \times \frac{\mu_{I'} - \mu_{O'}}{\sqrt{q \cdot \sigma_{O'}^2 + p \cdot \sigma_{I'}^2 + pq(\mu_{I'} - \mu_{O'})^2}} \end{aligned}$$

Therefore,

$$r_{GX} = \frac{\mu_{I'} - \mu_{O'}}{\sqrt{\frac{1}{p} \cdot \sigma_{O'}^2 + \frac{1}{q} \cdot \sigma_{I'}^2 + (\mu_{I'} - \mu_{O'})^2}}. \tag{13}$$

While $(\mu_{I'} - \mu_{O'})$ appears in both the numerator and the denominator of the formula for the point biserial correlation, a closer examination of Equation 13 reveals that $(\mu_{I'} - \mu_{O'})$'s influence on the numerator outweighs its influence on the denominator, as the latter is also affected by two positive variance terms (i.e., $\frac{1}{p} \cdot \sigma_{O'}^2 + \frac{1}{q} \cdot \sigma_{I'}^2$). The only exception where increase in $(\mu_{I'} - \mu_{O'})$ does not lead to increase in r_{GX} is when $(\frac{1}{p} \cdot \sigma_{O'}^2 + \frac{1}{q} \cdot \sigma_{I'}^2)$ equals zero (i.e., when $\sigma_{O'}^2$ and $\sigma_{I'}^2$ are both zero). In that particular scenario,

$$r_{GX} = \frac{\mu_{I'} - \mu_{O'}}{\sqrt{(\mu_{I'} - \mu_{O'})^2}} = 1.$$

(*Appendix continues*)

As long as attentive and IER responses vary within each condition (i.e., $\frac{1}{p} \cdot \sigma^2_{O'} + \frac{1}{q} \cdot \sigma^2_{I'} \neq 0$), the higher the mean difference on scores between IER and attentive groups, the greater the correlation between the dichotomous variable IER condition and observed scores, and the more confounded observed scores are by the presence of IER.

It should be noted that the presence of IER in survey data can be viewed as a special case of nonindependence of observations (Kenny & Judd, 1986). That is, common statistical techniques such as ordinary least square regression and ANOVA assume that responses are drawn from the same population, with observations independent from each other (Bliese & Hanges, 2004). Violation of the independence assumption can affect the Type I and Type II error rates in hypothesis testing (Bliese & Hanges, 2004). The critical condition of $M_{attentive} \neq M_{\mathrm{IER}}$ determines that IER and attentive data will have different means, resulting in two clusters on the survey data. The between-group variance stemming from the mean difference between the IER and attentive data thus confounds the true effect one is interested in testing.

Received August 28, 2013
Revision received October 16, 2014
Accepted November 11, 2014 ■