APPLIED PSYCHOLOGY

IAAP
THE INTERNATIONAL ASSOCIATION OF APPLIED PSYCHOLOGY

APPLIED PSYCHOLOGY: AN INTERNATIONAL REVIEW, 2016, *65* (2), 287–321
doi: 10.1111/apps.12058

Insufficient Effort Survey Responding: An Under-Appreciated Problem in Work and Organisational Health Psychology Research

Alyssa K. McGonagle*
Wayne State University, USA

Jason L. Huang
Michigan State University, USA

Benjamin M. Walsh
University of Illinois at Springfield, USA

Insufficient effort responding (IER) is problematic in that it can add a systematic source of variance for variables with average responses that depart from the scale midpoints. We present a rationale for why IER is of particular importance to Work and Organisational Health Psychology (WOHP) researchers. We also demonstrate its biasing effects using several variables of interest to WOHP researchers (perceived work ability, negative affectivity, perceived disability, work–safety tension, accident/injury frequencies, and experienced and instigated incivility) in two datasets. As expected, IER was significantly correlated with the focal study variables. We also found some evidence that hypothesised bivariate correlations between these variables were inflated when IER respondents were included. Corroborating IER's potential confounding role, we further found significant declines in the magnitude of the hypothesised bivariate correlations after partialling out IER. In addition, we found evidence for biasing (under-estimation) effects for predictors *not* contaminated by IER in multiple regression models where some predictors and the outcome were both contaminated by IER. We call for WOHP researchers to routinely discourage IER from occurring in their surveys, screen for IER prior to analyzing survey data, and establish a standard practice for handling IER cases.

* Address for correspondence: Alyssa K. McGonagle, Department of Psychology, Wayne State University, 5057 Woodward Ave, 7th Floor, Detroit, MI 48202, USA. Email: alyssa.mcgonagle@wayne.edu

INTRODUCTION

Survey research methods are commonly used in Work and Organisational Health Psychology (WOHP) research, as psychologists rely on individuals' reports to assess many constructs. For example, popular theoretical models of stress emphasise the importance of individual perceptions or appraisals of stress (Lazarus & Folkman, 1984), perceptions of personal resources (Hobfoll, 1989), and reports of emotions or affectivity, which are appropriately captured through self-report (e.g. Perrewé & Zellars, 1999; Spector, 1994). Indeed, self-report survey methods are even referred to as the "dominant method of data collection for research in occupational health psychology" (Eatough & Spector, 2013, p. 248).

For the most part, researchers assume that the data from these survey methods represent participants' evaluations and responses to the survey items, and that the data and conclusions drawn from them (for instance, about relations between substantive variables) are unbiased (i.e. free from systematic inflation or deflation due to factors unrelated to the focal variables). One exception pertains to concerns raised about bias (e.g. inflated relations) that may occur due to the presence of common method variance (CMV; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Another exception pertains to nonresponse bias, which may occur "when the individuals responding to a survey differ from nonrespondents on variables relevant to the survey topic" (Rogelberg & Luong, 1998, pp. 60–61). A substantial literature examines this issue (see Rogelberg & Stanton, 2007), and has identified predictors of nonresponse (e.g. Fauth, Hattrup, Mueller, & Roberts, 2013; Harman, Ellington, Surface, & Thompson, 2015; Mueller, Voelkle, & Hattrup, 2011; LaRose & Tsai, 2014). Importantly, similar to CMV, nonresponse does not necessarily mean that nonresponse bias will occur (Rogelberg & Luong, 1998).

Insufficient effort responding (IER)—which refers to a type of survey responding in which the individual does not pay attention to or read the items and/or the item instructions prior to responding to the item (Huang, Curran, Keeney, Poposki, & DeShon, 2012)—is another possible biasing factor and is the focus of the current study. Traditionally, IER is viewed as a source of measurement error, with evidence focused on IER's attenuating effect on bivariate relations via measurement error (e.g. Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; McGrath, Mitchell, Kim, & Hough, 2010). However, IER can, *under specific conditions*, systematically inflate correlations between variables affected by IER (Huang, Liu, & Bowling, 2015b). In this paper, we note that the conditions under which IER can inflate correlations are particularly relevant to WOHP researchers and replicate its inflating effects using variables common in WOHP research. Further, we demonstrate that predictors affected by IER can lead to systematic under-estimation of predictors unaffected by

IER in a multiple regression context where the outcome variable is affected by IER.

Despite its potential to spuriously inflate correlations between substantive variables, to date researchers generally have not paid sufficient attention to this issue. Liu, Bowling, Huang, and Kent (2013) surveyed Society for Industrial and Organizational Psychology members and results indicated that only 37 per cent reported checking for IER "a lot" or "almost all the time". Ran, Liu, Marchiondo, and Huang (2015) examined data-screening procedures pertaining to IER detection for articles published from 2012 to 2014 in the *Journal of Applied Psychology*, *Journal of Management*, and *Academy of Management Journal*, and found that IER screening procedures were reported in only three of 463 survey data samples.

Due to its potential for systematic bias inflating correlations, we argue that WOHP researchers should pay more attention to IER in their survey datasets. In the sections that follow, we explain the conditions under which IER is likely to systematically inflate correlations between study variables and discuss why this is an important problem for many WOHP constructs. We then demonstrate the effects of IER on substantive relations between common variables in WOHP research using two data samples. Moving beyond Huang et al.'s (2015b) study of IER's potential biasing effect on bivariate relations, we provide the first examination of its biasing effects in a multiple regression setting, particularly when some predictors are *unaffected* by IER. Throughout, we consider two different criteria for determining IER, which may be helpful for researchers deciding how to categorise workers as IER. Finally, we conclude with a discussion of recommended practices for researchers regarding screening for and addressing IER in their survey datasets.

IMPLICATIONS OF IER FOR WOHP RESEARCH

Huang et al. (2015b) state that, as a whole, scores for respondents engaging in IER tend to behave similarly to random responses, with resulting means on substantive measures centering around the scale midpoints.[1] According to Huang and colleagues, when scale means of study variables in attentive respondents hover around the scale midpoints, IER adds measurement error and attenuates bivariate relations (Spearman, 1904) and this is typically assumed in the literature (see McGrath et al., 2010). However, when scale means of study variables in attentive respondents depart from the scale midpoints, IER and attentive cases will then have different means (with IER means closer to the scale midpoints). This may lead to observed non-zero correlations between IER and the study variables. Across all survey measures in a given

[1] An underlying assumption is that these respondents are engaging in IER only and not in other response tendencies (e.g. socially desirable responding).

dataset, IER's correlation with a substantive measure depends negatively on the location of the substantive measures' means. That is, the lower a substantive variable's mean generally is, the more positive IER's correlation with that variable will be; the higher a substantive variable's mean is, the more negative IER's correlation with that variable will be. To the extent that substantive measures have means that depart from the scale midpoints, inflated correlations between study variables may occur (see Huang et al., 2015b, pp. 829–832, for a full summary of the mechanisms underlying this effect).

WOHP researchers frequently use survey research methods due to the perceptual nature of focal constructs such as stress appraisals, along with practical difficulties in obtaining "objective" or "other source" reports of constructs such as workplace incivility or discrimination. In addition, many variables of interest to WOHP researchers tend to have means that depart from scale midpoints. This is sometimes due to the nature of the scales used; many scales use frequency counts for low-frequency events. Examples include positively skewed variables such as counts of self-reported injuries and absences, experiences of negative emotions, and frequencies of experienced interpersonal workplace mistreatment. Variables that tend to be negatively skewed also exist, such as perceptions of work ability and physical health.

There are many studies using variables with means that depart from scale midpoints in the WOHP literature. For example, Jasperse, Ward, and Jose (2012) studied perceived religious discrimination and psychological symptoms in Muslim women using a survey. In this study, the mean of perceived discrimination was below the scale midpoint ($M = 1.97$, $SD = 0.90$) on a scale ranging from 1 (*rarely*) to 5 (*very often*) and the mean of psychological symptoms was 2.29 ($SD = 0.78$) on a scale ranging from 1 (*never*) to 5 (*very often*). This is just one example of many WOHP research articles that examine survey scales with average scores that depart from the scale midpoints. In a content analysis of research areas in the *Journal of Occupational Health Psychology* from 1996 to 2012, Piotrowski (2012) found that safety, accidents and injuries, physical health, and workplace violence were highly represented (ranging from 10 to 24 articles per topic) and worker mental health, incivility, absenteeism, sexual harassment and bullying were moderately represented, ranging from 7 to 8 articles per topic. To the extent that measures of these variables are affected by IER, spuriously inflated correlations among constructs may exist.

HYPOTHESISED EFFECTS OF IER ON EXAMPLES OF WOHP CONSTRUCTS AND INTER-CORRELATIONS

Now that we have discussed the likely implications of IER for WOHP research, our aim is to demonstrate its effects on relations between WOHP variables in three substantive topics of interest to WOHP researchers: perceived work ability, worker safety/accidents/injuries, and workplace incivility. We focused

on these three topics not only because they are frequently studied in the WOHP literature, but also because the key variables tend to have means that depart from scale midpoints, and thus are susceptible to IER's confounding influence. For each, we detail expected correlations among a set of substantive variables. In addition, we propose how we expect that IER will relate to these variables and how we expect the bivariate correlations between the substantive variables to be inflated by the presence of IER. First, however, given our rationale outlined above, we expect that IER will correlate positively with variables that have low means and negatively with variables that have high means, and that $r_{\text{IER-X}}$ is negatively correlated with the means of substantive measures when the means are rescaled (range from 0 to 1).

Hypothesis 1: The correlation of IER with study measures is negatively correlated with our measures' rescaled means.

Perceived Work Ability

First, we examine perceived work ability, which refers to a worker's perceived ability to continue working in his or her current job, given the challenges of the job and his or her mental and physical resources (Ilmarinen, Gould, Järvikoski, & Järvisalo, 2008). Work ability is a topic of substantial and growing interest in the Occupational Medicine/Health literatures; a search of the Thompson Reuters database for articles with the topic "work ability" in the Occupational Health, Ergonomics, Psychology and Management literatures revealed over 65 articles published in both 2013 and 2014. Recent examples in the WOHP literature include McGonagle, Fisher, Barnes-Farrell, and Grosch (2015) and Weigl, Müller, Hornung, Zacher, and Angerer (2013).

The construct of perceived work ability, when measured in working individuals, is subject to what is known as the "healthy worker effect" (e.g. Osmotherly & Attia, 2006; Sterling & Weinkam, 1986). That is, those who are currently working tend to generally perceive high levels of work ability. It is, therefore, perhaps not surprising that measures of self-reported perceived work ability in working populations tend to be skewed, with most participants choosing scores above the scale midpoints. For example, in a sample of 5,971 Finnish municipal employees ages 44–58, von Bonsdorff, Seitsamo, Ilmarinen, Nygård, von Bonsdorff, and Rantanen (2011) found that, of 2,710 men and 3,261 women, only 27 per cent and 20 per cent, respectively, scored below 7 on a scale of perceived work ability ranging from 0 to 10 (with 10 being the highest work ability score). Given that a typical response is above the scale midpoint, we expect IER individuals to generally score lower than non-IER individuals on measures of perceived work ability, and as a result, perceived work ability will be negatively correlated with IER. A similar argument can be made for

disability; individuals who are currently working are likely to score below the midpoint on a disability scale, on average. Therefore, we also expect that perceived disability will be positively correlated with IER.

Personal resources, such as sense of control and positive affectivity, correlate positively with individuals' perceptions of work ability, and in some cases, more strongly than job resources and job demands (McGonagle et al., 2015). Therefore, we also expect to see variables that indicate a lack of personal resources to be negatively correlated with perceived work ability. We examine negative affectivity, which we propose will positively relate to IER based on the rationale provided above related to its mean departing from scale midpoints. Further, we expect that perceived work ability will correlate negatively with perceived disability based findings from McGonagle et al. (2015) and von Bonsdorff et al. (2011). Due to the hypothesised presence of IER as a common confound, we propose that the correlations between perceived work ability and both negative affectivity and disability will be stronger (inflated) when IER respondents are included in the sample as opposed to when the IER respondents are removed.

Hypothesis 2a: Perceived work ability is negatively correlated with IER.
Hypothesis 2b: Negative affectivity is positively correlated with IER.
Hypothesis 2c: Disability is positively correlated with IER.
Hypothesis 2d: Perceived work ability is negatively correlated with negative affectivity.
Hypothesis 2e: Perceived work ability is negatively correlated with disability.
Hypothesis 3: The presence of IER inflates the observed relationships between (a) perceived work ability and negative affectivity and (b) perceived work ability and disability.

Worker Safety/Accidents/Injuries

Next, we examine variables related to worker safety and self-reported accidents and injuries on the job. Accidents and injuries have been of interest to psychologists and occupational health researchers and practitioners since the early 1900s (Kaplan & Tetrick, 2011). In the last 30 years, a focus on psychosocial work environment factors affecting safety, including safety climate, has dominated the research literature on safety. Safety climate research has burgeoned in recent years with 109 research articles published in 2013 alone, according to a Thompson Reuters search. Safety climate is a known leading indicator of safety behaviors and worker accidents/injuries (Christian, Bradley, Wallace, & Burke, 2009). Zohar (2010) states that "... safety climate represents the shared perceptions of the priority of safety compared to other competing priorities" (p. 1518). This is represented by work–safety tension, which is related to worker

self-reported safety behaviors (Morrow, McGonagle, Dove-Steinkamp, Walker, Marmet, & Barnes-Farrell, 2010).

Worker accidents and injuries are relatively infrequent in most worker samples, with low base rates and positively skewed distributions (see Zohar, 2000). We therefore expect to see scores generally below the scale midpoint for self-reported frequency of accidents/injuries and a positive correlation between self-reported accidents/injuries and IER. In addition, levels of work–safety tension—the tension felt when safety competes with other work priorities—tend to fall below the scale midpoint (e.g. McGonagle & Kath, 2010), particularly when examined across various jobs and industries, as physical safety hazards tend to be low across many job types. Further, we anticipate that the expected positive correlation between work–safety tension and accidents/injuries will be stronger when examining data that include IER respondents, compared with data in which IER respondents are removed.

Hypothesis 4a: Accidents/injuries are positively correlated with IER.
Hypothesis 4b: Work–safety tension is positively correlated with IER.
Hypothesis 4c: Work–safety tension is positively correlated with accidents/injuries.
Hypothesis 5: The presence of IER will inflate the observed relationship between work–safety tension and accidents/injuries.

Workplace Incivility

The third and final set of variables we examine relate to workplace incivility. Defined in Andersson and Pearson's (1999) seminal article, workplace incivility is a type of interpersonal workplace mistreatment that is conceptually distinguished from other types by its low intensity (Cortina & Magley, 2009). Uncivil behaviors include being put down, ignored, or addressed in an unprofessional manner at work (Cortina, Magley, Williams, & Langhout, 2001). Workplace incivility is a topic of considerable interest among WOHP scholars. Such interest spurred a qualitative review of the workplace incivility literature by Schilpzand, de Pater, and Erez (2014), who noted that "The past 15 years shows the 'take-off' trend of the new construct of workplace incivility." A Thompson Reuters search for articles on the topic of incivility in psychology and the behavioral sciences revealed over 120 published articles since 2012.

Measures of workplace incivility commonly used by WOHP researchers, such as the Workplace Incivility Scale (Cortina et al., 2001), capture the frequency with which respondents experience or engage in various rude behaviors. Given this frequency-based approach to measurement, distributions tend to be positively skewed with means that are lower than scale midpoints. For instance,

Gallus, Bunk, Matthews, Barnes-Farrell, and Magley (2014) report means of 0.76 and 0.62 for self-reported experienced incivility (i.e. victimisation) and instigated incivility (i.e. perpetration), respectively, on scales ranging from "0" (*never*) to "4" (*many times*). Because mean levels tend to be lower than scale midpoints, we expect that IER will positively relate to both experienced and instigated incivility.

We also examine the extent to which IER influences the associations between experienced and instigated incivility from and towards supervisors and co-workers.[2] Various theories (e.g. incivility spiral; Andersson & Pearson, 1999) and research (e.g. Gallus et al., 2014) show that experienced incivility is positively related to instigated incivility, so we also expected to see positive associations between the two constructs. Moreover, given the aforementioned influence of IER, we hypothesise that the positive association between experienced and instigated incivility will be stronger when IER cases are included as compared to when they are removed.

Hypothesis 6a: Experienced incivility is positively correlated with IER.
Hypothesis 6b: Instigated incivility is positively correlated with IER.
Hypothesis 6c: Experienced incivility is positively correlated with instigated incivility.
Hypothesis 7: The presence of IER will inflate the observed relationship between experienced incivility and instigated incivility.

IER'S EFFECTS ON PREDICTORS IN A MULTIPLE REGRESSION

In addition to examining the effects of IER on variables and their substantive inter-relations in a bivariate context, we sought to better understand the possible effects of IER in a multivariate context. Specifically, we investigate the influence of IER in a multiple regression setting wherein the outcome variable is contaminated by IER (i.e. correlated with IER) and the predictor variables range in their IER contamination. In general, we expect that predictors contaminated with IER (i.e. having a strong correlation with IER) will appear to have stronger regression weights when IER is present than absent, as IER adds a source of shared variance between these predictors and the outcome variable. Meanwhile, we expect that predictors not contaminated with IER (i.e. having near zero correlations with IER) will appear to have weaker regression weights when IER is present than absent, because the shared variance between the

[2] For the sake of simplicity, we focus only on within-source associations by examining the influence of IER on (a) the association between experienced incivility from supervisors and instigated incivility towards supervisors, and (b) experienced incivility from co-workers and instigated incivility towards co-workers.

predictors as a whole and the outcome variable will be attributed to IER-contaminated predictors to a greater extent.

Research Question: What are the effects of IER in a multiple regression context wherein the outcome variable is contaminated by IER and the predictor variables vary in their IER-contamination levels?

METHODS OF CAPTURING IER

There are several ways in which researchers can detect IER, including using post-hoc within-person correlations such as psychometric antonyms/synonyms, computing individual reliability coefficients across a large number of scales, identifying extreme long string response patterns, detecting an unlikely short amount of time that the participant spends on each page of an online survey, and embedding items within the survey to detect IER (see Huang et al., 2012). These methods have been found to largely converge in identifying IER (Huang et al., 2012; Huang et al., 2015b). For our demonstration, we detect IER using embedded survey items, including instructed response items (e.g. "Please select *strongly agree* for your response to this item"; see Meade & Craig, 2012) and infrequency items (e.g. "I have never used a computer"; see Huang, Bowling, Liu, & Li, 2015a, for a more detailed discussion of item types). We adopted these IER measures because, unlike post-hoc IER indices that are study-specific (e.g. psychometric antonyms/synonyms), embedded survey items generate scores that are more comparable across studies. Embedded survey items are also favorable in that researchers may easily include them in their survey data collections in an a priori manner.

We used two survey datasets to test our hypotheses. Dataset 1 included perceived work ability, disability, and negative affectivity and was used to test Hypotheses 2–3. Dataset 2 included safety climate, accidents/injuries, and incivility and was used to test Hypotheses 4–7. Both datasets included additional measures that were also used to test Hypothesis 1 (see Table 1).

METHOD

Dataset 1 Participants and Procedure

Dataset 1 was collected from working adults on Amazon.com's MTurk site as part of another study, and a subset of this sample is reported in the Appendix of McGonagle et al. (2015). Two thousand MTurk workers completed a prescreening survey that asked whether they were currently working for pay outside of MTurk and, if so, for how many hours per week. Of the 2,000, 569 individuals who reported working at least 20 hours per week for pay were

TABLE 1
Dataset 1 and Dataset 2 Additional Survey Measures, Response Scales, Coefficient Alphas, and Citations

Measure	*Scale*	*# of items*	α	*Citation*
Job Satisfaction[a]	1–5	3	.89	Cammann, Fichman, Jenkins, & Klesh (1983)
Mental Resources[a]	1–5	3	.82	Tuomi et al. (1998)
Life Satisfaction[a]	1–5	2	.68	Campbell, Converse, & Rogers (1976)
Employability[a]	1–5	5	.82	Berntson & Marklund (2007)
General Self-Efficacy[a]	1–5	8	.90	Chen, Gully, & Eden (2004)
Job Self-Efficacy[a]	1–5	8	.85	Chen, Goddard, & Casper (2004)
Anxiety[a]	0–4	6	.89	Derogatis & Melisaratos (1983)
Social Desirability[a]	0–1	10	NA*	Strahan & Gerbasi (1972)
Exhaustion Burnout[a]	1–5	3	.83	Demerouti, Mostert, & Bakker (2010)
Job Performance[a]	0–5	3	.88	Farh, Dobbins, & Cheng (1991)
Job Insecurity[a]	1–5	6	.84	Ashford, Lee, & Bobko (1989)
Conscientiousness[b]	1–5	4	.90	Goldberg (1999); Goldberg et al. (2006)
Neuroticism[b]	1–5	4	.85	Goldberg (1999); Goldberg et al. (2006)
Perfectionism[b]	1–7	10	.84	Hewitt & Flett (1991)
Co-worker Satisfaction[b]	0–3	6	.77	Brodke et al. (2009); Lake, Gopalkrishnan, Sliter, & Withrow (2010)
Sensitivity to Interpersonal Treatment[b]	1–7	4	.73	Bunk & Magley (2011)
Perceived Organisational Support[b]	1–7	3	.93	Eisenberger, Huntington, Hutchinson, & Sowa (1986)
Supervisor Satisfaction[b]	0–3	6	.81	Brodke et al. (2009); Lake, Gopalkrishnan, Sliter, & Withrow (2010)
Job Stress[b]	1–3	4	.77	Stanton, Balzer, Smith, Parra, & Ironson (2001)
Ethical Leadership[b]	1–7	10	.94	Brown, Treviño, & Harrison (2005)
Safety Leadership[b]	1–7	5	.94	Barling, Loughlin, & Kelloway (2002); Hayes, Perander, Smecko, & Trask (1998)
Civility Norms[b]	1–7	4	.92	Walsh et al. (2012)
Rumination–Interpersonal Offense[b]	1–5	6	.95	Wade, Vogel, Yu-Hsin Liao, & Goldman (2008)
Organisational Citizenship Behavior[b]	1–5	6	.88	Lee & Allen (2002)
Learning Goal Orientation[b]	1–7	4	.91	VandeWalle (1997)
Grit[b]	1–7	12	.73	Duckworth, Peterson, Matthews, & Kelly (2007)
Workaholism[b]	1–4	6	.68	Schaufeli, Taris, & Bakker (2008)

Note: [a] Indicates Dataset 1 and [b] indicates Dataset 2. *Index comprises the sum of endorsed items.

invited to complete a full survey, and 331 responded to the full survey and were paid $3.00.

Dataset 1 Measures

IER. We measured IER using six items that were embedded throughout the survey. Five were instructed response items (e.g. "Select strongly agree in response to this question") and one was an infrequency item ("There are 25 hours in a day"). For the instructed items, any response other than the instructed response represented failing the item and the respondent was marked as IER for that item. The infrequency item was embedded in a series of True/False questions; a "true" response indicated IER for that item. We computed a mean of the six items (Cronbach's $\alpha = .72$). We used two criteria to determine IER respondents and report our results based on both criteria in an effort to address complexities surrounding identification of IER. We first allowed one IER item to be failed based on recommendations of Huang et al. (2012) and the notion that failing one item can indicate transient measurement error due to imperfect scale reliability. In other words, individuals may inadvertently select an improper response option while still generally paying attention to survey items. We coded those who failed two or more IER items as "IER" ($n = 20$) and removed them. Others were coded as "attentive" ($n = 311$) and comprised "Sub-Sample A". Second, we created a "Sub-Sample B", in which 75 participants were identified as IER (due to their failing one or more IER items) and removed, leaving an attentive n of 256. In neither case did we count missing responses as IER.

Perceived Work Ability. We measured perceived work ability using a four-item scale adapted from Tuomi, Ilmarinen, Jahkola, Katajarinne, and Tulkki (1998) and validated in McGonagle et al. (2015; $\alpha = .83$). A sample item is, "Thinking about the mental demands of your job, how do you rate your current ability to meet those demands?" The response scale ranged from (0) *cannot currently work at all* to (10) w*ork ability at its lifetime best*.

Negative Affectivity. We measured negative affectivity using ten items from the Positive and Negative Affectivity Scales (Watson, Clark, & Tellegen, 1988; $\alpha = .92$). Participants were asked, "During the last 30 days, to what degree did you feel. . ." (e.g. "upset"). The response scale ranged from (0) *never* to (4) *almost always.*

Disability. We measured perceptions of disability using a 12-item scale ($\alpha = .94$) from the World Health Organization (2010), the WHO-DAS 2.0. Participants were asked in the last 30 days how much difficulty they had in 12

activities (e.g. "standing for long periods such as 30 minutes"). The response scale ranged from (0) *none* to (4) *extreme or cannot do*.

Dataset 2 Participants and Procedure

Dataset 2 was collected from working students and employees of a Midwestern university in the United States. An online survey was posted on the university's psychology participant pool website. Students who indicated that they worked at least 10 hours per week for pay were able to view the posting and sign up to participate for course credit. The survey was also posted to the university's internal website. University staff who worked at least 10 hours per week for pay completed the survey for a $10 gift card incentive. Initially, 546 individuals clicked on the link to take the survey, but 14 people who failed to complete more than half of the survey were removed from the dataset, resulting in an N of 532 for analysis.

Dataset 2 Measures

IER. Three infrequency items were adopted from Huang et al. (2015a): "I eat cement occasionally", "I can teleport across time and space", and "I have never used a computer". We created an IER scale ($\alpha = .74$) by calculating the mean of the coded items. Again, we used two criteria for classifying individuals as IER in Dataset 2. Responses were coded as IER if (a) the participant responded to the item (missing responses were not considered IER) and (b) they responded to the item with "neutral", "agree", "slightly agree", or "strongly agree" (any form of disagreement was not considered an IER response; see Huang et al., 2015a). In Sub-Sample A, 91 individuals who failed more than one IER item were coded as IER and removed (attentive $n = 441$). In Sub-Sample B, 170 individuals who failed one or more IER items were coded as IER and removed (attentive $n = 362$).

Experienced Incivility from Supervisors and Co-Workers. We measured source-specific incivility from supervisors ($\alpha = .93$) and co-workers ($\alpha = .94$) using separate, modified versions of the Workplace Incivility Scale (Cortina et al., 2001). Instructions stated to focus specifically on experiences from either supervisors or co-workers. A sample supervisor item is, "During the past year, have you been in a situation in which your [supervisor] . . . Made demeaning or derogatory remarks about you?" The response scales ranged from 1 (*never*) to 5 (*daily*).

Instigated Incivility Toward Supervisors and Co-Workers. We measured instigated incivility toward supervisors ($\alpha = .95$) and co-workers ($\alpha = .93$) separately using modified items from Blau and Andersson (2005). A sample supervisor item is, "During the last year, how often have you exhibited the following

behaviors towards your [supervisor] ... Paid little attention to a statement made by your supervisor or showed little interest in their opinion?" The response scales ranged from 1 (*never*) to 5 (*daily*).

Work–Safety Tension. We assessed work–safety tension with four items ($\alpha = .89$; see McGonagle, Walsh, Kath, & Morrow, 2014). Two of the items were modified from Zohar's (1980) effect of work pace on safety scale by Mueller, DaSilva, Townsend, & Tetrick (1999), one was from Dedobbeleer and Béland's (1991) measure of worker involvement in safety, and one was adapted from Mueller et al. (1999) by Hofmann and Mark (2006). A sample item is, "My job duties often interfere with my ability to act as safely as I would like." The response scale ranged from 1 (*strongly disagree*) to 7 (*strongly agree*).

Accidents/Injuries. Participants reported numbers of accidents and injuries at work using three items ($\alpha = .74$) from Smecko and Hayes (1999). Participants indicated how many minor injuries, major injuries, and near accidents they experienced at work in the past six months (ranging from 0 to 8 or more). A mean score was calculated to represent accidents/injuries.

TABLE 2
Dataset 1 Descriptive Statistics and Correlations with IER

					Lenient IER Cutoff		*Stringent IER Cutoff*	
Measure	$r_{IER\text{-}X}$	M_{ALL}	RM_{ALL}	SD_{ALL}	M_{ATT} N=311	M_{IER} N=20	M_{ATT} N=256	M_{IER} N=75
Work Ability	−.23***	7.95	0.80	1.62	8.04	6.66	8.07	7.55
Negative Affectivity	.28***	1.09	0.27	0.72	1.04	1.74	1.02	1.31
Disability	.38***	0.63	0.16	0.75	0.56	1.62	0.53	0.95
Job Satisfaction	−.04	3.83	0.71	0.94	3.84	3.61	3.85	3.73
Mental Resources	−.06	4.15	0.79	0.78	4.16	4.04	4.17	4.09
Life Satisfaction	.07	3.02	0.51	0.80	3.01	3.28	3.01	3.06
Employability	.01	3.40	0.60	0.57	3.40	3.48	3.40	3.42
General Self-Efficacy	−.14*	4.01	0.75	0.58	4.03	3.71	4.03	3.92
Job Self-Efficacy	−.15**	4.08	0.77	0.54	4.09	3.84	4.10	3.99
Anxiety	.26***	0.60	0.15	0.77	0.56	1.29	0.55	0.75
Social Desirability	−.08	6.40	0.64	1.46	6.42	6.05	6.45	6.23
Exhaustion Burnout	.03	2.77	0.44	1.02	2.77	2.72	2.77	2.76
Job Performance	−.06	4.12	0.78	0.65	4.14	3.85	4.13	4.12
Job Insecurity	.28***	1.56	0.14	0.65	1.52	2.13	1.50	1.73

Note: Full $N = 331$. * $p < .05$; ** $p < .01$; *** $p < .001$. M = Mean, RM = Rescaled Mean, M_{ALL} = Mean of full sample, M_{ATT} = Mean of attentive respondents, M_{IER} = Mean of IER respondents, and SD = Standard Deviation. $r_{IER\text{-}X}$ = correlation between IER scale score and a substantive measure. The first set of means was attained when IER was defined as participants missing two or more IER items; the second set was attained when IER was defined as participants missing one or more IER items.

RESULTS

Survey scale descriptive statistics and correlations with IER are in Tables 2 and 3.

Relations between $r_{IER\text{-}X}$ and Rescaled Means

As noted, the direction in which a substantive measure is correlated with IER, as denoted by $r_{IER\text{-}X}$, depends negatively on the mean of the substantive measure: IER should correlate positively with variables that have low means

TABLE 3
Dataset 2 Descriptive Statistics and Correlations with IER Measure

					Lenient IER Cutoff		*Stringent IER Cutoff*	
	$r_{IER\text{-}X}$	M_{All}	RM_{All}	SD	M_{ATT} $N=441$	M_{IER} $N=91$	M_{ATT} $N=362$	M_{IER} $N=170$
Experienced Incivility–Superv.	.30***	1.46	0.11	0.75	1.34	1.94	1.34	1.72
Experienced Incivility–Co-workers	.26***	1.55	0.14	0.82	1.46	1.97	1.44	1.78
Instigated Incivility–Supervisors	.46***	1.37	0.09	0.72	1.22	1.98	1.21	1.70
Instigated Incivility–Co-workers	.36***	1.43	0.11	0.75	1.31	1.61	1.30	1.70
Work–Safety Tension	.51***	2.35	0.23	1.34	2.07	3.72	1.98	3.16
Accidents/Injuries	.25***	1.96	0.12	1.41	1.82	2.62	1.79	2.34
Conscientiousness	−.20***	4.15	0.79	0.85	4.21	3.82	4.24	3.94
Neuroticism	.14**	2.96	0.49	0.98	2.90	3.26	2.91	3.07
Perfectionism	−.02	4.85	0.64	0.94	4.86	4.81	4.85	4.87
Co-worker Satisfaction	−.18***	2.03	0.68	0.87	2.09	1.70	2.11	1.83
Sensitivity to Inter. Treatment	−.13**	5.00	0.67	1.11	5.06	4.70	5.09	4.80
Perceived Org. Support	−.03	5.04	0.67	1.46	5.03	5.06	5.08	4.94
Supervisor Satisfaction	−.20***	2.11	0.70	0.92	2.20	1.69	2.20	1.91
Job Stress	−.03	1.41	0.47	1.10	1.44	1.28	1.41	1.42
Ethical Leadership	−.17***	5.10	0.68	1.25	5.18	4.67	5.24	4.79
Safety Leadership	−.11*	5.12	0.69	1.31	5.19	4.81	5.21	4.95
Civility Norms	−.11*	5.07	0.68	1.41	5.13	4.81	5.19	4.83
Rumination–Inter. Offense	.25***	1.94	0.23	0.97	1.83	2.45	1.78	2.26
Org. Citizenship Behavior	−.18***	3.76	0.69	0.77	3.82	3.48	3.84	3.59
Learning Goal Orientation	−.26***	5.46	0.74	1.13	5.59	4.81	5.62	5.13
Grit	−.07	4.64	0.61	0.74	4.68	4.48	4.65	4.63
Workaholism	.12**	2.42	0.47	0.57	2.39	2.57	2.40	2.48

Note: Full $N = 532$. * $p < .05$; ** $p < .01$; *** $p < .001$. M = Mean, RM = Rescaled Mean, M_{ALL} = Mean of full sample, M_{ATT} = Mean of attentive respondents, M_{IER} = Mean of IER respondents, and SD = Standard Deviation. $r_{IER\text{-}X}$ = correlation between IER scale score and a substantive measure. The first set of means was attained when IER was defined as participants missing two or more IER items; the second set was attained when IER was defined as participants missing one or more IER items.

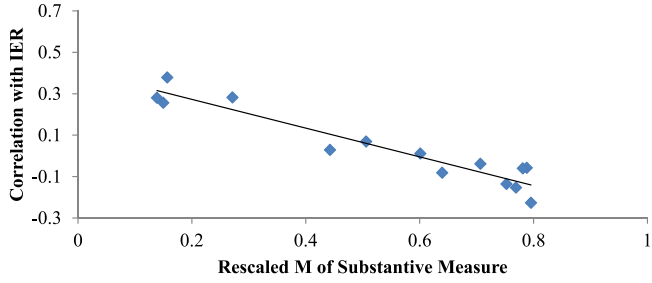

FIGURE 1. Scatterplot demonstrating that the correlation between IER and Dataset 1 study measures is negatively correlated with the overall rescaled scale means ($N = 14$; $r = -.95$).

and negatively with variables that have high means. We examined our datasets for the pattern of IER's associations with substantive measures (and replicated Huang et al., 2015b). Specifically, we expected to find that $r_{\text{IER-X}}$ is negatively correlated with substantive measures' rescaled means (rescaled to be between 0 and 1). We first computed the correlation $r_{\text{IER-X}}$ between IER and each study measure. We then rescaled variable means by making the lowest response category for each measure zero and the highest response category 1, with a midpoint of 0.5. Finally, we subjected the $r_{\text{IER-X}}$ values to Fisher's r to z transformation before correlating them with the corresponding rescaled variable means. The resulting correlations were $r = -.95$, $n = 14$, $p < .001$ for Dataset 1 and $-.93$, $n = 22$, $p < .001$ for Dataset 2 (see Figures 1 and 2 for scatterplots). These results support Hypothesis 1.

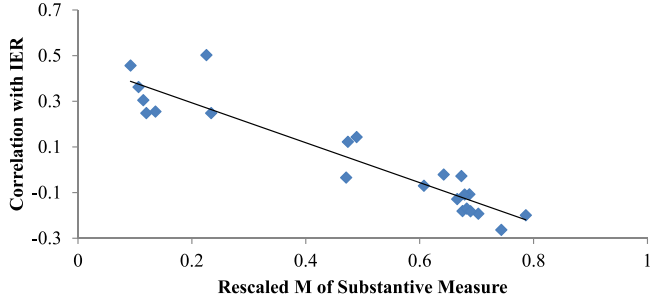

FIGURE 2. Scatterplot demonstrating that the correlation between IER and Dataset 2 study measures is negatively correlated with the overall rescaled scale means ($N = 22$; $r = -.93$).

Relations of Study Variables with IER and Differences between Correlations

As shown in Table 2, in Dataset 1 perceived work ability was negatively correlated with IER, $r = -.23$, $p < .001$. Thus, participants engaging in IER tended to report lower work ability than their attentive counterparts. Also, negative affectivity and perceived disability were positively correlated with IER, $r = .28$, $p < .001$ and $r = .38$, $p < .001$, respectively, indicating that participants engaging in IER tended to report higher levels of both negative affectivity and disability than attentive respondents, supporting Hypotheses 2a, 2b, and 2c. Further, perceived work ability was negatively correlated with negative affectivity ($r = -.13$, $p < .05$) and disability ($r = -.33$, $p < .001$), providing support for Hypotheses 2d and 2e.

Next, we examined differences in the substantive variable correlations between those calculated using the full Dataset 1 versus attentive respondents only (Sub-Sample A and Sub-Sample B). This presented a methodological challenge due to the nested nature of the samples (attentive datasets nested in and non-independent of the full dataset). As no existing options for comparing these correlations exist, we chose to use Fisher's r to z transformation to test the difference in correlations (Preacher, 2002) by treating them as independent correlations. Notably, such treatment made the current tests conservative for detecting significant differences between correlation coefficients because correlations obtained in nested samples are likely to be more similar than those from independent samples. Also, as noted by Kenny (1987), tests of differences in independent correlations generally have low power.

Results of correlation differences using both sub-samples are presented in Table 4. As noted above, the correlation between perceived work ability and negative affectivity in the full sample ($N = 331$) was $r = -.13$, $p < .05$, in Sub-Sample A ($n = 311$), it was $r = -.04$, $p > .05$, and in Sub-Sample B ($n = 256$) it was $r = -.09$, $p > .05$. Although one may draw dissimilar statistical conclusions about these bivariate relationships, the tests for differences between independent correlations were non-significant for both Sub-Sample A ($z = -1.14$, $p > .05$) and Sub-Sample B ($z = -0.48$, $p > .05$). Similarly, the differences in correlations were non-significant for the perceived work ability–disability relationship (Sub-Sample A $z = -1.10$, $p > .05$ and Sub-Sample B $z = -1.05$, $p > .05$). Therefore, we failed to find support for Hypotheses 3a and 3b.

As displayed in Table 3, both injuries and work–safety tension were positively correlated with IER in Dataset 2 ($r = .25$, $p < .001$ and $r = .51$, $p < .001$, respectively), indicating that those participants engaging in IER reported higher levels of both work–safety tension and workplace injuries and supporting Hypotheses 4a and 4b. As displayed in Table 4, work–safety tension was correlated with accidents/injuries ($r = .23$, $p < .001$) in the full sample, supporting Hypothesis 4c. The correlation using Sub-Sample A ($n = 441$) was $r = .20$,

TABLE 4
Differences between Correlations

	Full Sample	*Attentive Sub-Sample A*	*Attentive Sub-Sample B*	*Difference between Full Sample & A*	*Difference between Full Sample & B*
Variables	*r*	*r*	*r*	*z*	*z*
Work Ability – Negative Affectivity[a]	−.13*	−.04	−.09	−1.14	−0.48
Work Ability – Disability[a]	−.33***	−.25***	−.25***	−1.10	−1.05
Work–Safety Tension – Accidents/Injuries[b]	.23***	.20***	.16**	0.49	1.06
Supervisor Experienced – Instigated Incivility[b]	.61***	.47***	.43***	3.07**	3.63***
Co-worker Experienced – Instigated Incivility[b]	.70***	.63***	.59***	1.94*	2.76**

Note: [a] Indicates Dataset 1 (Full Sample $N = 331$; Sub-Sample A $N = 311$; Sub-Sample B $N = 256$). [b] Indicates Dataset 2 (Full Sample $N = 532$; Sub-Sample A $N = 441$; Sub-Sample B $N = 362$). Sub-Sample A is attentive respondents defined as those who are not missing two or more IER items. Sub-Sample B is attentive respondents defined as not missing one or more IER items. * $p < .05$; ** $p < .01$; *** $p < .001$. Differences between correlations were tested with tests of independent correlations using Fisher's *r* to *z* transformation.

$p < .001$, and in Sub-Sample B ($n = 362$) it was $r = .16$, $p < .01$. Yet, tests of the differences were again non-significant (Sub-Sample A $z = 0.49$, $p > .05$; Sub-Sample B $z = 1.06$, $p > .05$). Therefore, Hypothesis 5 was not supported.

Also in Dataset 2, experienced incivility from both supervisors and co-workers was positively correlated with IER as expected ($r = .30$, $p < .001$ and $r = .26$, $p < .001$, respectively), as was instigated incivility toward supervisors ($r = .46$, $p < .001$) and co-workers ($r = .36$, $p < .001$; see Table 3). These results indicate that participants engaging in IER reported higher levels of both experienced and instigated incivility. In addition, experienced incivility was correlated with instigated incivility for both supervisors ($r = .61$, $p < .001$) and co-workers ($r = .70$, $p < .001$). Hypotheses 6a–6c were supported. Further, the correlations between experienced supervisor incivility and instigated supervisor incivility in the subsets of Dataset 2 without IER were smaller in magnitude than those of the full sample as expected (Sub-Sample A $r = .47$ $p < .001$ and Sub-Sample B $r = .43$, $p < .001$). These differences were both significant (Sub-Sample A $z = 3.07$, $p < .01$ and Sub-Sample B $z = 3.63$, $p < .001$). The correlations between experienced co-worker incivility and instigated co-worker incivility were also smaller in magnitude than those calculated using the full sample (Sub-Sample A $r = .63$, $p < .001$ and Sub-Sample B $r = .59$, $p < .001$). These differences were both significant (Sub-Sample A $z = 1.94$, $p < .05$ and Sub-Sample B $z = 2.76$, $p < .01$). Hypothesis 7 was supported.

TABLE 5
Supplemental Analysis Results

Variables	*Zero-Order Correlation*	*Partial Correlation*	*Difference between Correlations*
Work Ability – Negative Affectivity[a]	−.13*	−.07	−.06**
Work Ability – Disability[a]	−.33***	−.27***	−.06**
Work–Safety Tension – Accidents/Injuries[b]	.23***	.12**	.11***
Supervisor Experienced – Instigated Incivility[b]	.61***	.56***	.05**
Co-worker Experienced – Instigated Incivility[b]	.70***	.68***	.03*

Note: Differences between correlations were tested using Olkin and Finn's (1995, Model C) test for full versus partial correlations. Partial correlations are with IER partialled out. [a] Indicates Dataset 1 and [b] indicates Dataset 2. * $p < .05$; ** $p < .01$; *** $p < .001$.

Supplemental Analyses of Correlation Differences

As noted, our tests of differences between the correlations were conservative in that they have low power (Kenny, 1987) and require an assumption of independence. Therefore, we examined results of an additional set of supplemental analyses in an effort to fully understand the potentially biasing influence of IER on substantive correlations. Specifically, we tested the differences between (a) bivariate correlations between substantive variables and (b) partial correlations in which IER was partialled out from both substantive variables. Results in Table 5 based on Olkin and Finn's (1995, Model C) test indicate significant reductions in effect size after partialling out IER in all cases. Thus, these supplemental analyses provide support for the notion that IER represents a source of common confound in the focal substantive correlations.

The Influence of IER in a Multiple Regression Context

Since most research in WOHP goes beyond bivariate relations among variables, it is important to understand the potential biasing effects of IER in a multivariate context. We assessed IER's influence on regression coefficients for predictors affected and unaffected by IER in a multiple regression. In both datasets, we regressed outcome variables affected by IER onto conceptually relevant predictor variables that ranged in the degree to which they were correlated with IER. Doing so allowed us to examine how IER might affect the manifested strengths of relations among predictors.

Using Dataset 1, we examined perceived disability as an outcome variable and perceived work ability and negative affectivity as predictors that are affected by IER as noted above. In addition, we added a third predictor, mental

TABLE 6A
Multiple Regression Results Predicting Perceived Disability (Dataset 1)

Predictor Variable	*Correlation of Predictor with IER*	*Full Sample β*	*Sub-Sample A β*	*Sub-Sample B β*
Work Ability	−.23***	−.18***	−.12*	−.07
Negative Affectivity	.28***	.36***	.32***	.26***
Mental Resources	−.06	−.28***	−.34***	−.41***
R^2		.37***	.35***	.35***

Note: All predictors entered simultaneously. Correlations were calculated using the full sample ($N = 331$). Sub-Sample A ($N = 311$) is attentive respondents who did not miss two or more IER items. Sub-Sample B ($N = 256$) is attentive respondents who did not miss one or more IER items. * $p < .05$; *** $p < .001$.

resources (Tuomi et al., 1998), as it was unaffected by IER (see Table 2), but is conceptually related to perceived disability. As displayed in Table 6a, the beta weights attained using the full sample were $\beta = -.18$, $p < .001$ for perceived work ability, $\beta = .36$, $p < .001$ for negative affectivity, and $\beta = -.28$, $p < .001$ for mental resources. When IER respondents were removed, the beta weights for perceived work ability and negative affectivity decreased, and the beta weight for mental resources increased. For Sub-Sample A, the beta weights were −.12, .32, and −.34, respectively, and for Sub-Sample B, the beta weights were −.07, .26, and −.41, respectively. In sum, the regression weights for mental resources (unaffected by IER) relating to disability (affected by IER) increased as IER participants were removed when other IER-contaminated predictors were included in the model.

We observed a similar pattern when examining results from Dataset 2. We first used accidents/injuries as the outcome and three predictors conceptually related to accidents/injuries: work–safety tension, safety leadership, and job stress. As displayed in Table 6b, work–safety tension was strongly affected by

TABLE 6B
Multiple Regression Results Predicting Accidents/Injuries (Dataset 2)

Predictor Variable	*Correlation of Predictor with IER*	*Full Sample β*	*Sub-Sample A β*	*Sub-Sample B β*
Work–Safety Tension	.51***	.20***	.15**	.10*
Safety Leadership	−.11***	−.18***	−.14**	−.17**
Job Stress	−.03	.08	.16**	.17**
R^2		.10***	.10***	.09***

Note: All predictors entered simultaneously. Correlations were calculated using the full sample ($N = 532$). Sub-Sample A ($N = 441$) is attentive respondents who did not miss two or more IER items. Sub-Sample B ($N = 362$) is attentive respondents who did not miss one or more IER items. * $p < .05$; ** $p < .01$; *** $p < .001$.

TABLE 6C
Multiple Regression Results Predicting Supervisor Incivility Instigation (Dataset 2)

Predictor Variable	*Correlation of Predictor with IER*	*Full Sample β*	*Sub-Sample A β*	*Sub-Sample B β*
Supervisor Incivility	.30***	.61***	.42***	.36***
Civility Norms	−.11*	−.02	−.12**	−.16**
R^2		.37***	.23***	.21***

Note: All predictors entered simultaneously. Correlations were calculated using the full sample ($N = 532$). Sub-Sample A ($N = 441$) is attentive respondents who did not miss two or more IER items. Sub-Sample B ($N = 362$) is attentive respondents who did not miss one or more IER items. * $p < .05$; ** $p < .01$; *** $p < .001$.

IER, safety leadership was weakly affected by IER, and job stress was unaffected by IER. In the full sample, both work–safety tension and safety leadership uniquely predicted accidents/injuries as expected ($\beta = .20$, $p < .001$ and $\beta = -.18$, $p < .001$, respectively), whereas job stress unexpectedly did not ($\beta = .08$, $p > .05$). Yet, when IER respondents were removed, the beta weights for work–safety tension and safety leadership generally decreased (Sub-Sample A $\beta = .15$, $p < .01$ and $\beta = -.14$, $p < .01$, respectively; Sub-Sample B $\beta = .10$, $p < .05$ and $\beta = -.17$, $p < .01$, respectively) and the beta weights for job stress increased and became significant ($\beta = .16$, $p < .01$ and $\beta = .17$, $p < .01$, respectively).

We found similar patterns when using incivility instigation as the outcome variable. We first examined incivility instigation toward a supervisor as predicted by experienced incivility from a supervisor (strongly related to IER) and civility norms (weakly related to IER; see Table 6c). When using the full sample, experienced incivility was a significant predictor ($\beta = .61$, $p < .001$); however, unexpectedly, civility norms was not ($\beta = -.02$, $p > .05$). Yet, when IER respondents were removed, the beta weight for experienced incivility decreased to $\beta = .42$ ($p < .001$) for Sub-Sample A and $\beta = .36$ ($p < .001$) for Sub-Sample B, and the beta weight for civility norms increased and became significant ($\beta = -.12$, $p < .01$ and $\beta = -.16$, $p < .01$ for the two sub-samples, respectively). When examining co-worker incivility instigation as the outcome and experienced incivility from co-workers (strongly affected by IER) and civility norms (weakly affected by IER) as predictors, we also saw a decrease in beta weights for experienced co-worker incivility when IER respondents were removed (from .71 to .59 and .56 for Sub-Samples A and B, respectively; see Table 6d). Beta weights for civility norms also increased (from .01 to −.07 and −.08 for Sub-Samples A and B, respectively); yet, civility norms did not become a significant predictor of co-worker incivility instigation when IER respondents were removed.

TABLE 6D
Multiple Regression Results Predicting Co-Worker Incivility Instigation (Dataset 2)

Predictor Variable	*Correlation of Predictor with IER*	*Full Sample β*	*Sub-Sample A β*	*Sub-Sample B β*
Co-worker Incivility	.26***	.71***	.59***	.56***
Civility Norms	−.11*	.01	−.07	−.08
R^2		.50***	.40***	.36***

Note: All predictors entered simultaneously. Correlations were calculated using the full sample ($N = 532$). Sub-Sample A ($N = 441$) is attentive respondents who did not miss two or more IER items. Sub-Sample B ($N = 362$) is attentive respondents who did not miss one or more IER items. * $p < .05$; *** $p < .001$.

DISCUSSION

Following Huang et al. (2015b), we challenge the prevalent view of the effect of IER as mere measurement error that attenuates observed relationships and put forth the view that IER is an issue that WOHP researchers should more consistently consider, given its potential to inflate correlations between variables with average scores that depart from the scale midpoints. We found evidence that IER is correlated with a number of WOHP constructs that fit those criteria across two datasets. As expected, respondents engaging in IER reported lower work ability and higher negative affectivity, disability, experienced and instigated incivility, work–safety tension, and accidents/injuries than attentive respondents. Further, we examined the impact of IER on five bivariate correlations between substantive variables affected by IER and found a stronger correlation in each case when IER individuals were included rather than excluded. Although our tests of the statistical difference between correlations based on the full sample versus attentive subsample provided mixed results, we did find consistent evidence that IER served as a common confound in bivariate correlations, as partialling out IER resulted in reduced correlations in all five tests. Importantly, we also extended existing knowledge about IER through our findings that simultaneously including IER-contaminated predictors alongside uncontaminated predictors in equations estimating IER-contaminated outcomes resulted in downwardly biased estimates of the regression coefficients for the uncontaminated predictor variables. Overall, our results suggest that if IER is left unchecked, skewed variables in the WOHP literature such as negative affectivity, workplace mistreatment, and perceived work ability may carry variance due to IER. Subsequently, bivariate correlations between such constructs may be inflated and uncontaminated predictors may be deflated when also included in a multiple regression context. Notably, although the focus of this study was on IER's effects on WOHP constructs, our findings have

relevance to any topic that relies on survey report data, particularly when the variables have scale means that depart from their midpoints.

It was not surprising that the comparison of substantive variable correlations in the full sample versus the attentive subsample reached statistical significance only sporadically. As explained earlier, we adopted a conservative test for the difference by treating the correlations as if they were independent. Notably, the removal of IER cases resulted in weaker correlations for five out of five bivariate relations, corresponding to a p value of .06 using a non-parametric sign test. This result is more consistent with the supplemental analyses that showed smaller correlations when IER was partialled out from substantive variables. The consistent reduction in the magnitude of correlations after removal of IER cases in two datasets indicates the possibility of a biased literature base for the associations among skewed variables if IER is not systematically screened from primary studies.

Our results imply that screening for IER can help address current measurement-related issues identified in Yang, Chang, and Lim (2014). Specifically, the authors note that "advances in measurement are needed . . . to differentiate overlapping WOHP constructs" (p. 564). Our findings suggest that removing IER substantially reduces correlations between experienced and instigated incivility, enabling WOHP researchers to better distinguish between the two constructs. This may be extended to other conceptually overlapping WOHP constructs as well and may enable more accurate tests of construct validity when examining variables with means that depart from scale midpoints. Relatedly, multicollinearity between two predictors both affected by IER may be reduced when removing IER respondents, enabling more accurate results of tests of the unique effects of such correlated predictors on outcome variables.

Extending current knowledge of IER, we also demonstrated that the presence of IER-contaminated predictors may cause downwardly biased estimates of predictors unaffected by IER in a multiple regression context when the outcome variable is also IER-contaminated.[3] This is likely because shared variance between predictors and the outcome variable is being attributed

[3] The reader may wonder about effects on outcome variables unaffected by IER. With our current focus, the logic for predictors unaffected by IER to be underestimated has to do with the inclusion of predictors affected by IER and outcomes affected by IER sharing variance associated with IER; when that variance is removed, a less biased estimate should be attained for predictors unaffected by IER. In contrast, when an outcome variable is not affected by IER, regression weights for predictors both affected and unaffected by IER will vary depending on their true relation with the outcome, the effects of IER on these predictors (i.e. systematic versus random), and the extent to which they are correlated. Therefore, we are unable to make any systematic predictions about changes in regression coefficients when the outcome variable is not contaminated with IER.

erroneously to IER-contaminated predictors. Thus, when IER participants are removed, a more accurate and stronger relation will be obtained between the unaffected predictors and the outcome variable. The notion that IER can also affect estimates of predictor variables unaffected by IER provides additional support for the notion that researchers should take steps to avoid IER and screen for it carefully in their samples.

Differential Effects of IER

Although we focused on instances in which IER may inflate bivariate associations (i.e. both variables have means away from scale midpoints), IER can also attenuate correlations in other instances. As is commonly assumed in the literature, the presence of IER can decrease measurement reliability for variables that have means near scale midpoints and, therefore, weaken observed correlations (see Huang et al., 2015b). Researchers may expect to see attenuation occur, therefore, when their variables have means that are close to the scale midpoints and when IER is likely to affect variable reliability—for instance, when a scale includes both positively and negatively scored items (as individuals who are not paying attention are likely to respond sub-optimally to reverse-keyed items). On the other hand, our results and those of Huang et al. (2015b) indicate that researchers should anticipate IER to have inflating effects on correlations when variable means depart from scale midpoints, as they do in variables discussed in this study. Importantly, whereas the attenuation scenario can increase Type II errors, the inflation scenario is likely to introduce Type I errors. Our results therefore provide impetus for WOHP researchers to guard against IER.

Rates and Implications of IER in Different Sampling Contexts

One might question the relevance of IER's influence in organisational samples, given the fact that our data were based on an online panel and a primarily student sample. We echo Landers and Behrend (2015) that sample types represent an artificial distinction, and argue that IER may occur in all kinds of samples. Indeed, instances of IER in organisational samples can be seen in both published journal articles (e.g. Hough et al., 1990; Stetz, Button, & Quist, 2012) and conference presentations (e.g. Nieminen, Kotrba, Denison, & Carter, 2014). Thus, we call for efforts to examine IER regardless of the sample source.

Rates of IER reported in the organisational literature vary. The rates of IER in this study are higher than those reported by Meade and Craig (2012), who reported 10–12 per cent. Some studies have observed higher rates (e.g. 26% removed by Cho & Allen, 2012, in an Mturk sample); yet, others have reported lower rates (e.g. 6% removed by Barber & Budnick, 2015, in a student sample). We may speculate about differences in IER rates based on sampling strategies.

An MTurk worker, for instance, may be motivated to provide better data quality than an undergraduate student who simply needs to submit a survey for course credit, due to the fact that MTurk workers rely on their online reputations to be qualified for certain assignments. We see some evidence supporting this notion based on the relative amounts of IER in our two datasets as noted above. Yet, this also likely depends on pay rates. We paid MTurk participants a fair amount in this study; it may be informative to compare rates of IER in our sample versus in samples where the pay is much lower. In addition, survey experience or savviness may also play a part in the detection of IER. MTurk respondents may be more experienced with survey responding than respondents from other sources (e.g. student sample); some may even anticipate quality assurance checks. Therefore, an important question for future research is whether explicit IER items (e.g. instructed response items) function as effectively as infrequency items in online paid samples. Further, as the number of studies reporting proportions of IER increase, meta-analytic techniques to determine relative amounts of IER across different sampling strategies (e.g. crowdsourcing, student samples, and organisational samples) would be informative.

Rates of IER also likely influence the extent to which IER is a potentially biasing problem. That is, the greater the proportion of the dataset that is IER, the greater likelihood that biased (inflated) results will occur, to the extent that the means of focal variables depart from the midpoints of the scales. We see some evidence for this in our study, as our proposed hypotheses are supported to a greater extent in Dataset 2, which had a relatively larger proportion of IER respondents than Dataset 1 (17% who missed two items or 32% who missed one item versus 6% who missed two items or 23% who missed one item in Dataset 1).

Cutoff Values for Determining IER

We examined two cutoff values for determining IER: failing two or more IER items (A) and failing one or more IER items (B). As may be expected, use of the more stringent criteria produced larger declines in bivariate correlations and larger changes in regression coefficients in multiple regressions. This may seem to imply that researchers should use more stringent criteria in determining IER participants. However, just like on any other psychological measure, an IER scale is unlikely to have perfect reliability and validity, and some attentive respondents may be flagged as IER due to transient errors (e.g. misunderstanding) on a particular IER item. Thus, the decision on a cutoff IER scale value is complex and should depend on several factors, such as the sample size and the degree to which variable means depart from scale midpoints.

Implications and Recommendations

Procedural Recommendations. We present a summary of recommendations in Table 7. In addition to detecting IER post-hoc, we echo Huang et al. (2012) in underscoring the importance of preventing IER from occurring. To this end, researchers may consider increasing rapport with their respondents by displaying a video clip of researchers reading the survey instructions (Ward, Meade, Gasperson, & Pond, 2014), issuing a benign note about the existence of IER detection tools (Huang et al., 2015a), and creating a tie between participant incentives and data quality (see Huang et al., 2015b). As noted, researchers have found that participants are more likely to provide comments when they have an interest in the topic (Harman et al., 2015)—interest may not be able to be manipulated by researchers, but providing some information on the survey topic and the projected use of the response data may help. Future research in this area should include interventions to reduce IER in online surveys, such as real-time feedback in online surveys for individuals who fail more than one IER item or go through survey pages in an unlikely rapid fashion.

Post-Hoc Remedies. Despite efforts to prevent IER from occurring via the procedural remedies noted, IER may still occur. Therefore, we recommend that WOHP researchers consistently screen for IER in their studies and remove individuals who exhibit high levels of IER. Specifically, we suggest that researchers include IER measures such as infrequency items (e.g. Huang et al., 2015a) and time measures (e.g. Huang, 2014) in their surveys, and set a priori cutoffs for detecting and removing IER respondents prior to data analysis

TABLE 7
Recommendations for Researchers

Deterring Strategies
1. Build rapport with respondents.
2. Issue a benign note about the existence of IER detection tools in the survey.
3. Create a tie between participant incentives and data quality.

Detecting IER
1. Plan to consistently screen for IER in surveys in work and organisational health psychology research studies, especially when studying constructs with means that tend to depart from scale midpoints.
2. Include a priori IER detection methods in surveys when possible (i.e. embedded IER items).
3. Set a priori cutoffs for detecting and removing IER cases prior to data analysis.

Reporting Results
1. Provide details on the IER measure used, including the source, example items, and reliability/validity evidence if available.
2. Note the IER cutoff used and the basis for its selection.
3. Document the number of respondents determined to exhibit IER.

based on the considerations noted above. When infrequency items are used, we recommend the use of multiple items to measure IER. This enables researchers to check for the reliability of their IER measures. However, there may be cases when only a single item is used to measure IER, and researchers are unsure of its reliability. In these cases, it may be especially advisable to report results with and without detected IER cases. When reporting results on the full sample, researchers who have constructs with average scores that depart from scale midpoints should check for potential biasing effects of IER on relations between their study variables. When removing anyone failing the single IER item from their analysis, researchers should also recognise that the potential change in results may stem from true influence due to IER or from measurement error on that single IER item.

We acknowledge that researchers may encounter difficulties adding IER questions to organisational surveys. Organisational sponsors may not immediately see a benefit of including them and may view them as wasting employees' time. However, the results of our paper, together with Huang et al. (2015b), provide exactly the reason for including some form of IER measure in organisational settings: Results may indicate that two variables are related when in fact the relationship is spurious due to the presence of IER. Equipped with the present results, researchers should be ready to describe the benefits of being able to detect IER with the use of only a few extra survey items to organisational practitioners; the benefits include being able to better trust responses attained. It is arguably more of a waste of time to conduct a survey for which the data obtained are of questionable quality. Researchers may also weigh the potential benefits with the cost when proposing their use in organisational settings. Contexts where IER is more likely include when survey participation is mandated and when the survey is anonymous (Meade & Craig, 2012). In these cases, researchers may want to advocate more strongly for inclusion of IER items.

Reporting. As noted, organisational researchers generally do not report treatment of IER in published studies (Ran et al., 2015). This may reflect an overall inattention to the issue of IER (see Liu et al., 2013), or simply reflect researchers' omission in reporting such screening. In either case, the non-treatment of IER cases and the non-reporting of such treatment can hinder WOHP researchers' ability to replicate previous findings (Simons, 2014). We offer the following recommendations for reporting: inclusion of items used to measure IER and internal consistency reliability of the measure (e.g. coefficient alpha), description of the cutoff used to determine IER respondents and rationale for the chosen cutoff, and numbers of respondents who were determined to exhibit IER and removed from the study.

Additional Future Research Directions

One possible implication of the biasing effects of IER in WOHP research with scales departing from the midpoints is the tendency to control for negative affect and negative affectivity in stress research (e.g. Spector, Zapf, Chen, & Frese, 2000). To the extent that the substantive measure and negative affectivity are both correlated with IER—and negative affectivity typically is, as scores tend to fall below the scale midpoint—the correlations between substantive measures and negative affectivity may be inflated. It is likely when WOHP researchers control for negative affectivity that they are also partialling out some variance due to IER. When first eliminating IER cases, we would expect to see weakened correlations between negative affectivity and the substantive variables. As a result, the role of negative affectivity as a control variable may decrease when IER respondents are removed. Another related implication is that, when including negative affectivity in multiple regression equations alongside uncontaminated variables, failing to remove IER respondents may result in deflated regression coefficients for the uncontaminated variables. Both scenarios indicate the need for researchers to prevent IER from occurring and screening for and removing IER individuals from their datasets. Interestingly, we found evidence to support the notion that the presence of IER makes negative variables appear worse than they likely actually are—stressors appear more common and workers appear less healthy. Future research may consider comparing survey responses from IER and attentive samples with comparable measures attained externally—e.g. health or disability reports from a doctor or reports of injuries from organisational records.

Given our results, researchers may wonder what the effects of IER would be on indirect effects in the context of mediation analyses, as well as in moderation analyses. We also expect IER to affect theoretically relevant indirect effects (mediations), yet to a lesser extent because change to a product term of two slope coefficients would be smaller, compared to its constituent paths. For example, if path a is inflated from .20 to .30, and path b is inflated from $-.20$ to $-.30$, the product term ab is only inflated from $-.04$ to $-.09$.[4] The effects of IER on statistical moderations are likely

[4] We examined theoretically relevant mediation models in the two datasets and found relatively small changes to indirect effects when examining the products of two slopes that are both individually affected by IER. For example, we examined an indirect effect of job self-efficacy to work disability via perceived work ability using Sample 1. The indirect effect with the full sample ($n = 331$) was ab $= -.09$, $p = .02$, and the indirect effect with the attentive subsample ($n = 311$) was ab $= -.04$, $p = .06$. Using Sample 2, we found that the indirect effect of experienced co-worker incivility on co-worker-targeted incivility perpetration via co-worker satisfaction was ab $= .035$, $p = .002$ with the full sample ($n = 530$) and ab $= .044$, $p = .005$ with the attentive subsample ($n = 441$).

complex and would require the use of simulated data to understand. Future research should more completely examine IER's effects on both mediation and moderation analyses.

Another important future research direction for IER involves attempts to better understand IER behaviors. We assume that individuals exhibiting IER do not pay attention/read survey items and/or response scales in an effort to respond quickly. Yet, there may be different "degrees" of IER wherein individuals range from simply clicking down the page using mostly the same response number/option or randomly, to spending some time/paying some attention to and reading each item, but not sufficient enough, leading them to potentially misinterpret some statements or inadvertently click the wrong response option.

As mentioned, the literature on survey nonresponse may be informative for understanding IER. In particular, IER could potentially be conceptualised as a form of response bias, as IER respondents may be systematically different from attentive respondents. In both areas, there is a need to understand the motivation behind IER/providing sufficient effort, just as there is a need to understand motivation to respond/not respond (Rogelberg & Stanton, 2007). Indeed, research on correlates of response rates could shed light on reasons for IER. For instance, Fauth et al. (2013) found that group-level job satisfaction was positively correlated with group-level response rates across multiple samples (see also Mueller et al., 2011; Taris & Schreurs, 2007). In other words, more satisfied groups are more likely to respond to organisational surveys. Also, Harman et al. (2015) found that participants in a training program were more likely to provide comments (training reactions) when they had an interest in the topic. We speculate that this could extend to individuals having an interest in the survey topic being more likely to pay attention to items and display lower levels of IER. Further, it is possible that incentives may affect IER similarly to how they affect nonresponse (LaRose & Tsai, 2014). Finally, although we interpret IER's relationships with substantive variables as evidence that IER is a methodological nuisance, it is possible that the relations between IER and some study variables reflect, to some extent, a substantive personality influence underlying IER. For example, it is possible that engaging in IER is a manifestation of low conscientiousness and that low conscientiousness predisposes one to experience workplace accidents (see Hypothesis 4a). Future research is warranted and these propositions need to be tested empirically. A better understanding of the underlying motivation behind IER behavior can be obtained by real-time verbal protocols and follow-up interviews. Researchers may also develop more sophisticated detection methods to discern different patterns in which an individual may engage in IER, and subsequently assess their impacts on observed relationships.

Conclusion

In conclusion, IER is under-appreciated by WOHP researchers, yet is demonstrated to have potential systematic biasing effects on substantive variable correlations in survey research when average responses to variables depart from scale midpoints. We recommend that researchers aim to prevent IER from occurring in their research studies, and also routinely screen for and remove IER cases prior to analysis and report their treatment of IER in publications.

REFERENCES

Andersson, L.M., & Pearson, C.M. (1999). Tit for tat? The spiraling effect of incivility in the workplace. *Academy of Management Review*, *24*, 452–471. doi: 10.5465/AMR.1999.220213

Ashford, S.J., Lee, C., & Bobko, P. (1989). Content, causes, and consequences of job insecurity: A theory-based measure and substantive test. *Academy of Management Journal*, *32*(4), 803–829.

Barber, L.K., & Budnick, C.J. (2015). Turning molehills into mountains: Sleepiness increases workplace interpretive bias. *Journal of Organizational Behavior*, *36*, 360–381. doi: 10.1002/job.1992

Barling, J., Loughlin, C., & Kelloway, E.K. (2002). Development and test of a model linking safety-specific transformational leadership and occupational safety. *Journal of Applied Psychology*, *87*, 488–496.

Berntson, E., & Marklund, S. (2007). The relationship between perceived employability and subsequent health. *Work & Stress*, *21*, 279–292.

Blau, G., & Andersson, L. (2005). Testing a measure of instigated workplace incivility. *Journal of Occupational and Organizational Psychology*, *78*, 595–614.

Brodke, M.R., Sliter, M.T., Balzer, W.K., Gillespie, J.Z., Gillespie, M.A., Gopalkrishnan, P., Lake, C.J., Oyer, B., Withrow, S., & Yankelevich, M. (2009). *The Job Descriptive Index and Job in General (2009 Revision): Quick reference guide*. Bowling Green, OH: Bowling Green State University.

Brown, M.E., Treviño, L.K., & Harrison, D.A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, *97*, 117–134.

Bunk, J.A., & Magley, V.J. (2011). Sensitivity to interpersonal treatment in the workplace: Scale development and initial validation. *Journal of Occupational and Organizational Psychology*, *84*, 395–402.

Cammann, C., Fichman, M. Jenkins, G.D., Jr., & Klesh, J.R. (1983). Assessing the attitudes and perceptions of organizational members. In S.E. Seashore, E.E. Lawler III, P.H. Mirvis, & C. Cammann (Eds.), *Assessing organizational change: A guide to methods, measures and practices* (pp. 71–139). New York: John Wiley & Sons.

Campbell, A., Converse, P.E., & Rodgers, W. (1976). *The quality of American life: Perceptions, evaluations, and satisfactions*. New York: Russell Sage Foundation.

Chen, G., Goddard, T.G., & Casper, W.J. (2004). Examination of the relationships among general and work-specific self-evaluations, work-related control beliefs, and job attitudes. *Applied Psychology: An International Review*, *53*, 349–370.

Chen, G., Gully, S.M., & Eden, D. (2004). General self-efficacy and self-esteem: Toward theoretical and empirical distinction between correlated self-evaluations. *Journal of Organizational Behavior*, *25*, 375–395.

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

Christian, M.S., Bradley, J.C., Wallace, J.C., & Burke, M.J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, *94*, 1103–1127.

Cortina, L.M., & Magley, V.J. (2009). Patterns and profiles of response to incivility in the workplace. *Journal of Occupational Health Psychology*, *14*, 272–288. doi: 10.1037/a0014934

Cortina, L.M., Magley, V.J., Williams, J.H., & Langhout, R.D. (2001). Incivility in the workplace: Incidence and impact. *Journal of Occupational Health Psychology*, *6*, 64–80.

Dedobbeleer, N., & Béland, F. (1991). A safety climate measure for construction sites. *Journal of Safety Research*, *22*, 97–103. doi: 10.1016/0022-4375(91)90017-P

Demerouti, E., Mostert, K., & Bakker, A.B. (2010). Burnout and work engagement: A thorough investigation of the independency of both constructs. *Journal of Occupational Health Psychology*, *15*, 209–222.

Derogatis, L., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine*, *13*, 595–605.

Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D.R. (2007). Grit: Perseverence and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*, 1087–1101.

Eatough, E.M., & Spector, P.E. (2013). Quantitative self-report methods in occupational health psychology research. In M. Wang, R. Sinclair, & L. Tetrick (Eds.), *Research methods in occupational health psychology* (pp. 248–267). Boca Raton, FL: Taylor & Francis.

Eisenberger, R., Huntington, R., Hutchinson, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology*, *71*, 500–507.

Gallus, J.A., Bunk, J.A., Matthews, R.A., Barnes-Farrell, J.L., & Magley, V.J. (2014). An eye for an eye? Exploring the relationship between workplace incivility experiences and perpetration. *Journal of Occupational Health Psychology*, *19*, 143–154. doi:10.1037/a0035931

Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe, Vol. 7* (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., & Gough, H.C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96.

Farh, J.L., Dobbins, G.H., & Cheng, B.S. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and US workers. *Personnel Psychology*, *44*, 129–147.

Fauth, T., Hattrup, K., Mueller, K., & Roberts, B. (2013). Nonresponse in employee attitude surveys: A group-level analysis. *Journal of Business and Psychology*, *28*, 1–16.

Harman, R.P., Ellington, J.K., Surface, E.A., & Thompson, L.F. (2015). Exploring qualitative training reactions: Individual and contextual influences on trainee commenting. *Journal of Applied Psychology*, *100*, 894–916.

Hayes, B.E., Perander, J., Smecko, T., & Trask, J. (1998). Measuring perceptions of workplace safety: Development and validation of the Work Safety Scale. *Journal of Safety Research*, *29*, 145–161.

Hewitt, P.L., & Flett, G.L. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology*, *60*, 456–470.

Hobfoll, S.E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, *44*, 513–524. doi: 10.1037/0003-066X.44.3.513

Hofmann, D.A., & Mark, B.A. (2006). An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Personnel Psychology*, *59*, 847–869. doi: 10.1111/j.1744-6570.2006.00056.x

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. doi: 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* (1276), 1–8. doi: 10.3389/fpsyg.2014.01276

Huang, J.L., Bowling, N.A., Liu, M., & Li, Y. (2015a). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology*, *30*, 299–311. doi: 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 responding to surveys. *Journal of Business and Psychology*, *27*, 99–114. doi: 10.1007/s10869-011-9231-8

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

Ilmarinen, J., Gould, R., Järvikoski, A., & Järvisalo, J. (2008) Diversity of work ability. In R. Gould, J. Ilmarinen, J. Järvisalo, & S. Koskinen (Eds.), *Dimensions of work ability: Results of the Health 2000 Survey* (pp. 13–24). Helsinki: Finnish Institute of Occupational Health.

Jasperse, M., Ward, C., & Jose, P.E. (2012). Identity, perceived religious discrimination, and psychological well-being in Muslim immigrant women. *Applied Psychology: An International Review*, *61*, 250–271. doi: 10.1111/j.1464-0597.2011.00467.x

Kaplan, S., & Tetrick, L.E. (2011). Workplace safety and accidents: An industrial and organizational psychology perspective. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (pp. 455–472). Washington, DC: American Psychological Association.

Kenny, D.A. (1987). Testing measures of association. In D.A. Kenny (Ed.), *Statistics for the social and behavioral sciences* (pp. 270–291). Boston, MA: Little, Brown.

Lake, C.J., Gopalkrishnan, P., Sliter, M.T., & Withrow, S. (2010). The Job Descriptive Index: Newly updated and available for download. *The Industrial-Organizational Psychologist*, *48*(1), 47–49.

Landers, R.N., & Behrend, T.S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *8*(2), 142–164.

LaRose, R., & Tsai, H.Y.S. (2014). Completion rates and non-response error in online surveys: Comparing sweepstakes and pre-paid cash incentives in studies of online behavior. *Computers in Human Behavior*, *34*, 110–119.

Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

Lee, K., & Allen, N.J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *Journal of Applied Psychology*, *87*, 131–142.

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*, *51*(1), 32–38.

McGonagle, A.K., Fisher, G.G., Barnes-Farrell, J.L., & Grosch, J.W. (2015). Individual and work factors related to perceived work ability and labor force outcomes. *Journal of Applied Psychology*, *100*, 376–398. doi: 10.1037/a0037974

McGonagle, A.K., & Kath, L. (2010). Work–safety tension, perceived risk, and worker accidents: A meso-mediational model. *Journal of Safety Research*, *41*, 475–479. doi: 10.1016/j.jsr.2010.09.002

McGonagle, A.K., Walsh, B.M., Kath, L.M., & Morrow, S.L. (2014). Civility norms, safety climate, and safety outcomes: A preliminary investigation. *Journal of Occupational Health Psychology*, *19*, 437–452. doi: 10.1037/a0037110

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. doi: 10.1037/a0019216

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

Morrow, S.L., McGonagle, A.K., Dove-Steinkamp, M.L., Walker, C.T., Jr., Marmet, M., & Barnes-Farrell, J.L. (2010). Relationships between psychological safety climate facets and safety behavior in the rail industry: A dominance analysis. *Accident Analysis and Prevention*, *42*, 1460–1467. doi: 10.1016/j.aap.2009.08.011

Mueller, L., DaSilva, N., Townsend, J., & Tetrick, L. (1999, April). An empirical evaluation of competing safety climate measurement models. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Mueller, K., Voelkle, M.C., & Hattrup, K. (2011). On the relationship between job satisfaction and non-response in employee attitude surveys: A longitudinal field study. *Journal of Occupational and Organizational Psychology*, *84*, 780–798.

Nieminen, L., Kotrba, L., Denison, D., & Carter, N. (2014, May). Insufficient effort responding: Uncovering an insidious threat to data quality. In J.L. Huang

& M. Liu (Co-chairs), Insufficient effort responding to surveys: From impact to solutions. Symposium presented at the annual conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

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

Osmotherly, P., & Attia, J. (2006). The healthy worker survivor effect in a study of neck muscle performance measures in call-centre operators. *Work*, *26*, 399–406.

Perrewé, P.L., & Zellars, K.L. (1999). An examination of attributions and emotions in the transactional approach to the organizational stress process. *Journal of Organizational Behavior*, *20*, 739–752.

Piotrowski, C. (2012). Occupational health psychology: Neglected areas of research. *Journal of Instructional Psychology*, *39*, 189–191.

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. doi: 10.1037/0021-9010.88.5.879

Podsakoff, P.M., MacKenzie, S.B., & Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 1–31.

Preacher, K.J. (2002, May). Calculation for the test of the difference between two independent correlation coefficients [Computer software]. Available from: http://quantpsy.org.

Ran, S., Liu, M., Marchiondo, L.A., & Huang, J.L. (2015). Difference in response effort across sample types: Perception or reality? *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *8*, 202–208.

Rogelberg, S.G., & Luong, A. (1998). Nonresponse to mailed surveys: A review and guide. *Current Directions in Psychological Science*, *7*(2), 60–65.

Rogelberg, S.G., & Stanton, J.M. (2007). Introduction: Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods*, *10*, 195–209.

Schaufeli, W.B., Taris, T.W., & Bakker, A.B. (2008). It takes two to tango: Workaholism is working excessively and working compulsively. In R.J. Burke & C.L. Cooper (Eds.), *The long work hours culture: Causes, consequences and choices* (pp. 203–226). Bingley, UK: Emerald.

Schilpzand, P., de Pater, I.E., & Erez, A. (2014). Workplace incivility: A review of the literature and agenda for future research. *Journal of Organizational Behavior.* Advance online publication. doi: 10.1002/job.1976

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

Smecko, T., & Hayes, B. (1999, April). Measuring compliance with safety behaviors at work. Paper presented at the 14th annual conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201–292. doi: 10.2307/1412107

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.

Spector, P.E., Zapf, D., Chen, P.Y., & Frese, M. (2000). Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior*, *21*, 79–95.

Stanton, J.M., Balzer, W.K., Smith, P.C., Parra, L.F., & Ironson, G. (2001). A general measure of work stress: The Stress in General Scale. *Educational and Psychological Measurement*, *61*, 866–888.

Sterling, T.D., & Weinkam, J.J. (1986). Extent, persistence and constancy of the healthy worker or healthy person effect by all and selected causes of death. *Journal of Occupational Medicine*, *28*, 348–353.

Stetz, T.A., Button, S.B., & Quist, J. (2012). Rethinking carelessness on job analysis surveys: Not all questions are created equal. *Journal of Personnel Psychology*, *11*, 103–106. doi: 10.1027/1866-5888/a000061

Strahan, R., & Gerbasi, K.C. (1972). Short, homogeneous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *28*, 191–193.

Taris, T.W., & Schreurs, P.J. (2007). How may nonresponse affect findings in organizational surveys? The tendency-to-the-positive effect. *International Journal of Stress Management*, *14*, 249–259.

Tuomi, K., Ilmarinen, J.A., Jahkola, A., Katajarinne, L., & Tulkki, A. (1998). *Work Ability Index*, 2nd edn. Helsinki: Finnish Institute of Occupational Health.

VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement*, *8*, 995–1015.

von Bonsdorff, M.B., Seitsamo, J., Ilmarinen, J., Nygård, C., von Bonsdorff, M.E., & Rantanen, T. (2011). Work ability in midlife as a predictor of mortality and disability in later life: A 28-year prospective follow-up study. *Canadian Medical Association Journal*, *183*, E235–E242.

Wade, N.G., Vogel, D.L., Liao, K.Y.H., & Goldman, D.B. (2008). Measuring state-specific rumination: Development of the Rumination about an Interpersonal Offense Scale. *Journal of Counseling Psychology*, *55*, 419–426.

Walsh, B.M., Magley, V.J., Reeves, D.W., Davies-Schrils, K.A., Marmet, M., & Gallus, J.A. (2012). Assessing workgroup norms for civility: The development of the Civility Norms Questionnaire-Brief. *Journal of Business and Psychology*, *27*, 407–420. doi: 10.1007/s10869-011-9251-4

Ward, M.K., Meade, A.W., Gasperson, S., & Pond, S.B. (2014, May). Manipulating instructions to reduce careless responding on Internet-based surveys. Paper presented at the 28th annual meeting of the Society for Industrial and Organizational Psychology, Honolulu, HI.

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.

Weigl, M., Müller, A., Hornung, S., Zacher, H., & Angerer, P. (2013). The moderating effects of job control and selection, optimization, and compensation strategies on the age–work ability relationship. *Journal of Organizational Behavior*, *34*, 607–628.

Williams, L.J., & Anderson, S.E. (1994). An alternative approach to method effects by using latent-variable models: Applications in organizational behavior research. *Journal of Applied Psychology*, *79*, 323–331.

World Health Organization (2010). Measuring Health and Disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0), edited by T.B. Üstün, N. Kostanjsek, S. Chatterji, & J. Rehm. Accessed from: http://whqlibdoc.who.int/publications/2010/9789241547598_eng.pdf

Yang, L.-Q., Chang, C.-H.D., & Lim, V.K.G. (2014). Call for papers: New methods in work and organizational health psychology. *Applied Psychology: An International Review*, *63*, 564–566. doi: 10.1111/apps.12027

Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology*, *65*, 96–102. doi: 10.1037/0021-9010.65.1.96

Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology*, *85*, 587–596.

Zohar, D. (2003). The influence of leadership and climate on occupational health and safety. In D.A. Hofmann & L.E. Tetrick (Eds.), *Health and safety in organizations: A multilevel perspective* (pp. 201–230). San Francisco, CA: Jossey-Bass.

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis and Prevention*, *42*, 1517–1522. doi: 10.1016/j.aap.2009.12.019