Insufficient Effort Responding to Surveys as a Threat to Validity: The Perceptions and Practices of SIOP Members

Mengqiao Liu
Wayne State University

Nathan A. Bowling
Wright State University

Jason L. Huang
Wayne State University

Tiana A. Kent
Wayne State University

To properly test a theoretical model or provide empirically driven solutions to organizational problems, I-O researchers must collect *valid* data. Needless to say, maximizing validity is an explicit goal of research methodology, and very few SIOP members would dispute the importance of validity. It is possible, however, that even when researchers use measures that are *generally* valid, a subset of participants may still fail to provide valid data. This possibility is reflected in several mature research areas including social desirability (Ones, Viswesvaran, & Reiss, 1996), faking (Mueller-Hanson, Heggestad, & Thornton, 2003), and the "nuisance" effects of trait negative affectivity (Spector, Zapf, Chen, & Frese, 2000). We believe that insufficient effort responding (IER; Huang, Curran, Keeney, Poposki, & DeShon, 2012)—a largely overlooked behavior and the focus of this paper—poses yet an additional threat to the validity of self-report data.

What Is IER?

IER, "a response set in which the respondent answers a survey measure with low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses" (p. 100, Huang et al., 2012), can manifest itself in different ways. In some cases, participants may engage in IER by responding randomly to a series of questionnaire items. A given participant, for example, may choose "strongly agree" for the item "I am satisfied with my job" and carelessly choose "disagree" for the nearly identical item "I am happy with my job." IER may also result in a more systematic response pattern. A participant displaying this form of IER, for instance, may choose the "slightly agree" option for 30 consecutive items.

The effects of IER on research findings are potentially very serious. IER—particularly when it is manifested as random responding—would likely lower the reliability of psychological measures (Huang et al., 2012; Meade & Craig, 2012). Imagine, for instance, a research participant who randomly responds to a series of 20 self-report Conscientiousness items. The inclusion of that participant's data would likely lower the sample's reliability estimate for the Conscientiousness scale. Perhaps even more alarming, the presence of IER can in some instances inflate the relationships between conceptually distinct constructs (Huang, Bowling, & Liu, 2013) and thus increase the probability of a Type-I error. In other instances it can attenuate the relationships between conceptually related constructs (Liu, Huang, Bowling, & Bragg, 2013) and result in greater probability of a Type-II error. Translating the research findings into an applied example, the presence of IER may lead to an adoption of an invalid selection test (Type-I error) or a rejection of a valid selection test (Type-II error).

Critics may counter that IER is too rare to pose a serious threat to the validity of research data, but this position is not supported by recent research findings, which suggest that approximately 10% to 12% of research participants display evidence of IER (Meade & Craig, 2012). Empirical tests confirm that IER generally occurs at a rate sufficient to negatively impact the validity of self-report measures (Huang et al., 2013; Liu et al., 2013). Given its prevalence and its potential effects, it is not surprising that researchers have developed several strategies for detecting IER.

Correspondence concerning this paper should be addressed to Mengqiao Liu, Department of Psychology, Wayne State University, 5057 Woodward Ave, Detroit, MI 48202 USA; mengqiao.liu@wayne.edu

Strategies for Detecting IER

Below we discuss five strategies for detecting IER cases for potential removal from further analyses.

Infrequency approach. In the infrequency approach, participants are asked to answer self-report items that include a response option that can be safely assumed as the correct response. For example, one can assume that "no" is the accurate response to the question "I eat cement occasionally" (Huang et al., 2013). The presence of IER is inferred when participants provide responses that deviate from the assumed correct response.

Repeated-item approach. In the repeated-item approach, participants are asked to respond to the same survey item multiple times. Participants, for example, could be asked to respond to the item "I am satisfied with my organization" near the beginning of the questionnaire and then again later in the questionnaire. IER is assumed to be present when a given participant provides contradictory responses to an item that appears multiple times.

Inconsistency approach. The inconsistency approach uses correlations or similarity scores to examine within-person relationships between items that are highly related among participants in general. The item "I am satisfied with my supervisor," for example, should be positively related to the item "I am happy with my supervisor," and both should be negatively related to the item "I am dissatisfied with my supervisor." Within-person relationships that deviate from the above pattern would indicate IER.

Response pattern approach. The response pattern approach focuses on the detection of suspicious response patterns across a series of items. A participant may engage in IER, for instance, by selecting "neither agree nor disagree" as his or her response to 50 consecutive items. Another participant may respond in a regular pattern, such as a participant who alternates between providing "strongly disagree" and "strongly agree" responses. It is unlikely that a careful survey respondent would display such response patterns.

Response time approach. In the response time approach, IER is assumed to have occurred when a given participant completes a questionnaire unusually quickly relative to other participants. For example, it is likely that a particular participant who completed a questionnaire in 5 minutes has displayed IER if the average participant requires 30 minutes to complete the same questionnaire. The use of the response time approach is facilitated by electronic questionnaires, which typically allow researchers to record when a participant has begun and completed the questionnaire.

To What Extent Do I-O Researchers Recognize IER as a Threat to Validity?

We suspect that I-O researchers generally have limited knowledge of IER and as a result they typically eschew the use of methods designed to detect and minimize the occurrence IER. First, unlike other potential threats to validity (e.g., social desirability, trait negative affectivity), IER is seldom discussed as a potential methodological problem in I-O graduate training and in commonly used textbooks.[1] Although research has been conducted to examine other sources of measurement error, such as respondent fatigue, we argue that IER may pose a broader scope of problems beyond fatigue. Specifically, fatigue occurs because participants are *unable* to engage in careful responding, whereas IER may occur when participants can respond carefully but choose not to. In addition, fatigue-related issues tend to be more relevant in longer surveys where respondents may engage in IER-like behaviors toward the middle or end of the survey (Baer, Ballenger, Berry, & Wetter, 1997; Berry et al., 1991). IER, on the other hand, can stem from other factors such as respondents' lack of interest in the survey, limited contact between the researcher and the participants, and environmental distractions (Meade & Craig, 2012). Under such circumstances, IER may be present throughout a survey and should be identified accordingly. Furthermore, the practice of screening for IER is rarely reported in journal articles, with limited methodological studies devoted to the detection of IER.

It is thus important to document I-O researchers' perceptions and practices concerning IER. In this study we surveyed SIOP members to examine: (a) the extent to which IER is perceived as a significant problem, (b) the extent to which researcher detect IER with various strategies, and (c) the extent to which researchers believe they have found evidence of IER in their own data.

---

[1] Indeed, a cursory examination of six measurement and testing textbooks showed that only one (Furr & Bacharach, 2013, *Psychometrics: An Introduction*) directly mentions careless responding briefly as a source of response bias.

Method

Participants

An invitation email was sent to 2,360 professional SIOP members using e-mail addresses obtained via the SIOP Community Directory. Members interested in participating were directed to an online questionnaire. The sample consisted of 254 SIOP members (39% women, mean age = 43 years). Respondents self-identified as scientists (37%), practitioners (23%), or both scientists and practitioners (40%). Approximately 67% of participants held academic positions.

Measures

We designed a short survey to investigate SIOP members' beliefs about IER and their use of methods designed to detect IER. To minimize survey length, each study variable was assessed with a single-item measure.

IER-specific practices. We used four items to examine common practices related to IER. Specifically, we measured respondents' perceived impact of IER (*perceived impact*), effort in dealing with IER (*effort*), and the perceived frequency of IER in surveys (*perceived frequency*; see Table 1). Responses were made on a five-point Likert scale (1 = *not at all*; 5 = *almost all the time/extremely*). A fourth item, *detection approaches*, assesses the extent to which respondents utilize the following approaches: (a) infrequency approach, (b) repeated-item approach, (c) inconsistency approach, (d) response pattern approach, and (e) response time approach. To assist participants in answering the *detection approaches* item, we provided them with descriptions of each approach (see Table 2).

General survey practices. We used four items to assess respondents' general survey practices, including data collection method, typical amount of data collected, types of samples used, and types of student samples used (see Table 1). Specifically, for *data collection method*, we asked respondents to indicate the percentage of times they utilize (a) online surveys, (b) paper-and-pencil surveys, (c) surveys in lab setting, and (d) experiments. For *typical amount of data collection*, we asked respondents to estimate the sample size of the datasets they survey in a year. *Types of sample* assessed the percentage of times respondents' studies involve (a) student sample, (b) organizational sample, and (c) online paid participant sample (e.g., Mechanical Turk). For the *types of student sample*, respondents indicated if the research involves mostly undergraduate students, graduate students, or both.

Table 1
Items on IER and Survey Practices

Practices	Item
Perceived impact	To what extend do you think insufficient effort responding impacts your survey findings?
Effect	How often do you make an effort to deal with insufficient effort responding?
Perceived frequency	How often do you find IER in your surveys?
Detection approaches	What are the techniques that you use for screening and dealing with insufficient effort responding?
Data collection method	How often do you use each of the following data collection methods?
Typical amount of data collection	What is the typical amount of data collection you do in a calendar year?
Types of sample	What type of sample does your research usually involve?
Types of student sample	If your research involve a student sample, which of the following types are most representative?

*Demographic information***.** We asked participants to report their age, gender, and professional role (i.e., scientist, practitioner, or scientist/practitioner).

Results

What Are SIOP Members' General Perceptions of IER?

On average, respondents reported moderate levels of perceived impact of IER, M = 2.62 (SD = .81), effort in dealing with IER, M = 3.13 (SD = 1.34), and perceived frequency of IER in survey studies, M = 2.60 (SD = 1.19), with a plurality of respondents answering "A little" for perceived impact and effort (52% and 33%, respectively) and "Sometimes" for perceived frequency (43%; See Figure 1 through 3)

Table 2
Description of the Detection Approaches

Detection approach	Description
Infrequency approach	Uses items on which all or virtually all honest and attentive participants should provide the same response (e.g., “I was born on February 31st”).
Repeated-item approach	Asks participants to respond to a particular item multiple times within the survey.
Inconsistency approach	Examines within-person correlation or similarity in responses to items that are highly correlated.
Response pattern approach	Examine if a participant provides a suspicious pattern of responses (e.g., a long string of the same response options).
Response time approach	Screen for overly fast survey completion times.

Overall, there was a significant difference among the detection approaches respondents adopted, $F(5, 1509) = 28.37$, $p < .01$. Post hoc analyses revealed that response pattern approach ($M = 41.41\%$) was the most commonly used technique, followed by response time approach ($M = 21.49\%$), inconsistency approach ($M = 16.16\%$), infrequency approach ($M = 15.25\%$), and repeated-item approach ($M = 14.81\%$). No significant differences were detected among the frequency of use of these latter four techniques. In addition to the aforementioned approaches, respondents identified additional techniques to screen and deal with IER, such as screening out respondents with excessive missing values, using instructional manipulation checks, checking for univariate/multivariate outliers, and using open-ended questions. Several respondents also indicated methods to motivate survey respondents, such as "direct plea for engaged participation," offering feedback to respondents on their survey results, and training and better communication.

We also examined intercorrelations among IER-specific beliefs and practices. Results revealed that perceived impact was significantly associated with both effort, $r = .21$, $p < .01$, and perceived frequency, $r = .38$, $p < .01$. That is, SIOP members are more likely to exert effort to deal with IER when they also believe that the perceived impact of IER is high. As expected, effort was also positively correlated with perceived frequency, $r = .47$, $p < .01$, indicating that IER is more frequently found among those who make an effort to look for it. In addition, higher levels of perceived frequency were associated with higher frequencies in utilizing infrequency approach, response pattern approach, and response time approach, indicating that the use of these specific approaches may lead to identifying more IER compared to the other approaches (see Table 3).

Effects of Sample Types

On average, organizational samples (58%) are most frequently used in research among members of SIOP, followed by student samples (27%), and online paid samples (8%). Interestingly, sample types were shown to be related to IER beliefs and practices (see Table 3). In particular, the more frequent use of organizational samples was negatively associated with perceived impact, $r = -.21$, $p < .01$, effort, $r = -.22$. $p < .01$, and perceived frequency, $r = -.28$, $p < .01$, whereas the frequent use of paid online samples positively correlated with perceived impact, $r = .14$, $p < .05$, effort, $r = .31$, $p < .01$, and perceived frequency, $r = .33$, $p < .01$. In addition, the use of student samples was positively related with perceived frequency, $r = .18$, $p < .01$. In sum, these findings suggests that researchers may be least skeptical of data quality when organizational samples rather than paid online or student samples are used.

Effects of Data Collection Methods

Descriptive statistics revealed that online surveys (66%) were the most popular method for data collection, followed by paper- and-pencil surveys (18%), while the other methods were utilized less than 7% of the time. We used correlations to examine whether data collection methods were associated with IER beliefs and practices. In general, SIOP members' methods of data collection were not associated with their IER perceptions and practices (*r*s ranged from -.11 to .15).

Differences Across Professional Roles

ANOVA revealed a significant overall difference in effort, $F(2, 247) = 4.56$, $p < .05$, depending upon one’s professional role. Post hoc analyses indicated that both scientists ($d = .48$, $p < .05$) and scientist –practitioners ($d = .43$, $p < .05$) reported higher effort than did practitioners. In addition, perceived frequency differed across roles, $F(2, 222) = 5.74$, $p < .01$, such that scientists ($d = .62$, $p < .01$) as well as scientist –practitioners ($d = .48$, $p < .05$) reported higher perceived fre-

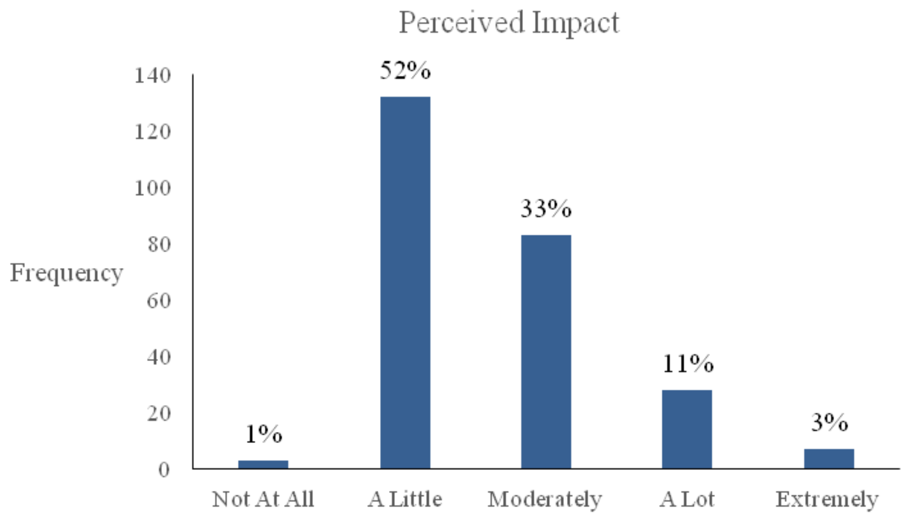

Figure 1: **Distribution of Responses on Perceived Impact**

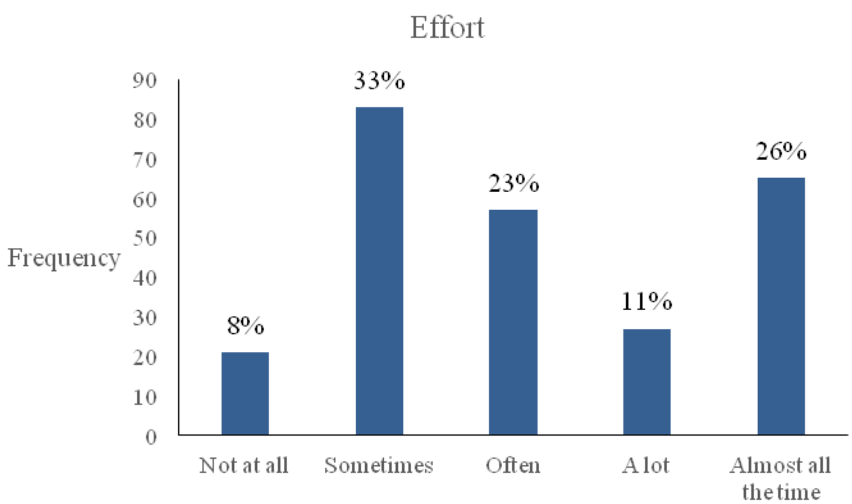

Figure 2: **Distribution of Responses on Effort**

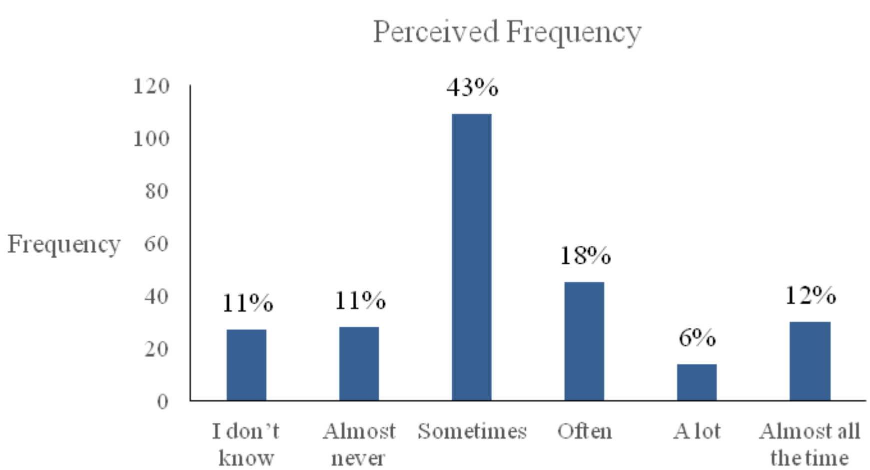

Figure 3: **Distribution of Responses on Perceived Frequency**

Table 3
Descriptive Statistics and Correlations for IER Practice

	1	2	3	4	5	6	7	8	9	10	11
1. Perceived impact											
2. Effort	0.21										
3. Perceived frequency	0.38	0.47									
4. Infrequency	0.12	0.31	0.29								
5. Repeated-item	0.10	0.04	-0.12	0.22							
6. Inconsistency	0.11	0.27	0.12	0.23	0.15						
7. Response pattern	0.06	0.43	0.25	0.20	-0.03	0.22					
8. Response time	0.15	0.33	0.34	0.13	0.06	0.14	0.39				
9. Student samples	0.12	0.07	0.18	0.18	0.11	-0.02	0.13	0.15			
10. Organizational samples	-0.21	-0.22	-0.28	-0.20	-0.03	-0.09	-0.18	-0.25	-0.70		
11. Online paid samples	0.14	0.31	0.33	0.23	-0.11	0.12	0.16	0.29	0.02	-0.41	
M	2.62	3.13	2.60	15.25	14.81	16.16	41.41	21.49	26.56	57.74	8.35
SD	0.81	1.34	1.19	31.12	28.47	29.91	40.00	33.75	32.40	39.51	18.59

Note . N = 253, except for correlations involving perceived frequency, where N = 226, as "I don't know, because I don't look for IER" was coded as missing. When $|r| > .13$, $p < .05$; when $|r| > .17$, $p < .01$; when $|r|$? .22, $p < .001$.

quency than did practitioners. No significant difference, however, was found between scientists and scientists–practitioners in either effort or perceived frequency. Perceived impact also did not differ significantly across roles.

Given the negative association between the use of organizational samples and effort, we suspected that the difference in effort across roles could be attributed to their primary sample types. Thus, we conducted ANCOVA controlling for organizational samples. Results showed that when controlling for the amount of organizational samples used, the significant effects of role on effort disappeared, $F(2, 245) = 1.30$, *ns*. In other words, the discrepancies in effort across roles were likely driven by the amount of organizational samples they use. Therefore, practitioners may not be making as much effort in dealing with IER because they mostly use organizational samples and perceive such data source as more valid than other sources (i.e., student samples and online paid samples).

Discussion

Despite growing evidence that IER is a prevalent problem and that it has unwanted effects on scale validity (Huang et al., 2013; Liu et al., 2013), SIOP members generally appear to consider it a minor or moderate issue and hence often do little to circumvent its effects. The current findings might in fact underestimate the extent to which SIOP members have overlooked the problem of IER, given that SIOP members who perceive IER as a threat may be more likely to have responded to our questionnaire. Although the positive associations among perceived impact, effort, and perceived frequency appear self-evident, it raises a serious issue: If researchers/practitioners turn a blind eye to IER, they are far less likely to identify IER in their survey results, thus creating a vicious circle that perpetuates inaction on IER. This phenomenon echoes Kepes and McDaniel’s (in press) criticism on a lack of a methods-related belief system that concerns the procedures of measurement, data collection, and analysis (see LeBel & Peters, 2011) such that the ignorance toward IER may pose yet another threat to the robustness of results in I-O research. We believe this finding is disconcerting and hope to call for a better understanding of IER building upon this article as well as recent empirical findings (e.g., Huang et al., 2012; Meade & Craig, 2012) and to educate researchers and practitioners about the undesirable effects of IER.

Interestingly, we found that IER was perceived to be a lesser threat among those who use organizational samples as opposed to paid online samples (e.g., Mechanical Turk) or student samples, which in turn lead to less effort in dealing with IER among nonacademic SIOP member than among academic members. These findings are consistent with the tacit assumption that IER is most prevalent within student samples and paid online samples and within datasets gathered using electronic questionnaires. On the other hand, organizational samples may be perceived as more motivated to pay attention in survey studies due to researchers’ active involvement in data collection or due to their perceptions of the research as being more closely related to their own interests. Despite what participants in the current study assume; however, we note evidence that IER is a potential threat in applied organizational research (e.g., Calsyn & Winter, 1999; Green & Stutzman, 1986; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990). In an organizational setting, employees may be too busy or too distracted with work activities to give sufficient effort in survey responding. Without assessing IER, practitioners will have no knowledge of the extent to which IER may have been present in the samples and thus no control of IER's detrimental effect in the development and validation of a scale, the assessment of a training program, or the interpretation of an organizational survey. Furthermore, the impact of IER has on organizational data may depend on the purpose of the survey, such that IER may become more problematic in situations where employees are less motivated (e.g., filling out a survey for training development) versus situations where the incentives are relatively high (e.g., selection). As a result, future research may further examine the quality of data collected from organizational samples and explore techniques that enable survey researchers to better detect and deal with IER across different settings. Although participants in organizational surveys may largely avoid engaging in IER—particularly when the survey addresses personally relevant content—some organizational practices (such as the "oversurveying" of an organization's workforce) may in fact encourage IER.

Call for Commentary

1. Do you think IER is an important issue to data quality? Why or why not?
2. What are the strategies that you use to screen for IER?
3. What is the biggest obstacle(s) that keeps you from looking for or examining IER?
4. Under what conditions do you think IER is most likely to occur?
5. Do you think some survey participants are predisposed to habitually engage in IER?
6. What influence do you think IER may have on your study results?
7. Do you think there has been sufficient coverage of the issue of IER in our graduate training?

References

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

Berry, D. T. R., Wetter, M. W., Baer, R. A., Widiger, T. A., Sumpter, J. C., Reynolds, S. K., & Hallam, R. A. (1991). Detection of random responding on the MMPI-2: Utility of F, back F, and VRIN scales. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 418–423.

Calsyn, R. J., & Winter, J. P. (1999). Understanding and controlling response bias in needs assessment studies. *Evaluation Review*, *23*, 399-417.

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An Introduction.* Thousand Oaks, CA: Sage.

Green, S. B., & Stutzman, T. (1986). An evaluation of methods to select respondents to structured job-analysis questionnaires. *Personnel Psychology*, *39*, 543-564.

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.

Huang, J. L., Bowling, N. A., & Liu, M. (2013). *Insufficient effort responding: Clarifying measurement construct validity via the use of a detection scale*. Manuscript submitted for publication.

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.

Kepes, S., & McDaniel, M. A. (in press). How trustworthy is the scientific literature in I-O psychology? *Industrial and Organizational Psychology: Perspectives on Science and Practice.*

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology, 15*, 371-379.

Liu, M., Huang, J. L., Bowling, N. A., & Bragg, C. (2013, April). *Attenuating effect of insufficient effort responding on relationships between measures*. Paper presented at the 28th Annual Conference for the Society for Industrial and Organizational Psychology, Houston, TX.

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

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C., III (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel decisions: The red herring. *Journal of Applied Psychology, 81*, 660- 679.

Smith, P. C., & Stanton, J. M. (1998). Perspectives on the measurement of job attitudes: The long view. *Human Resources Management Review, 8*, 367-386.

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.