SIOP White Paper Series

SOCIETY for INDUSTRIAL and ORGANIZATIONAL PSYCHOLOGY

SCIENCE FOR A SMARTER WORKPLACE

Insufficient Effort Responding to Psychological Assessments: Practical Advice for Combatting a Serious Threat to Data Quality

Nathan A. Bowling, Jason L. Huang, and Justin A. DeSimone

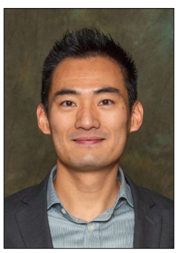
A White Paper prepared by the Visibility Committee of the Society for Industrial and Organizational Psychology.
6635 Central Ave., #105, Toledo, OH 43617

Table of Contents

Authors

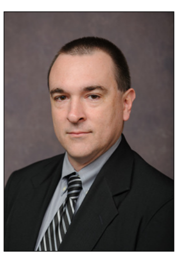
Nathan A. Bowling, PhD
University of Central Florida

Nathan Bowling is a professor of Psychology at the University of Central Florida, where he serves as the associate director of the I-O Psychology PhD Program. His research interests include occupational stress, counterproductive work behavior, and research participant carelessness. Nathan has authored 70 peer-reviewed articles, many of which have appeared in top-tier journals, including the *Journal of Applied Psychology, the Journal of Personality and Social Psychology,* and *Organizational Research Methods.* His research was featured in Joireman and Van Lange's (2015) American Psychological Association book *How to Publish High-Quality Research*, he was included in Aguinis et al.'s (2017) list of the top 2% of most cited authors in I-O psychology textbooks, and he was listed in Ioannidis et al.'s (2020) database of the world's most cited scientists. Nathan is a fellow of two professional organizations—SIOP and APS. As a PI, Nathan has been awarded over $1M in external funding since 2021. He is (as of 2025) the incoming editor-in-chief of the *International Journal of Stress Management.*



Jason L. Huang
Michigan State University.

Jason L. Huang, PhD, is a professor in the School of Human Resources and Labor Relations at Michigan State University, where he also serves as its director. His research focuses on individual differences, employee adaptability, training and transfer, and response effort. Jason's work has been published in top-tier journals such as *Journal of Applied Psychology, Journal of Management, Personnel Psychology, Psychological Bulletin, Journal of Personality and Social Psychology*, and *Organizational Research Methods.* Jason is a Fellow of both SIOP and the Association for Psychological Science, chair-elect of the Research Methods Division of the Academy of Management, and a former associate editor for *Journal of Management.*



Justin A. DeSimone
University of Alabama.

Justin A. DeSimone is an associate professor of Management and the Cooper Endowed Faculty Excellence Fellow at the University of Alabama. His research focuses on innovative measurement techniques as well as issues related to research methodology, psychometrics, and statistics. Justin's work has appeared in some of the top journals in the field, including the *Journal of Management, Organizational Research Methods,* and the *Journal of Applied Psychology*. Justin is currently serving as an associate editor at *Organizational Research Methods* and was previously an associate editor at the *Journal of Organizational Behavior*. He also served on the editorial review board for these journals, as well as the *Journal of Management* and *Journal of Business and Psychology.* Justin is a regular contributor to the Consortium for the Advancement of Research Methods and Analysis (CARMA), also serving as an advisory board member and associate editor for short courses. Additionally, Justin serves on the executive committee of the Research Methods Division of the Academy of Management.

Insufficient effort responding (IER) on psychological assessments is a well-studied topic with far-reaching implications that stretch across many organizational domains (Huang et al., 2012; Maniaci & Rogge, 2014; Meade & Craig, 2012).[1] IER, unfortunately, threatens data quality in assessments of psychological and organizational topics. Researchers have, for instance, detected IER in various assessments, including low-stakes employee surveys (e.g., Bowling et al., 2016, Study 1), job analysis questionnaires (e.g., Morgeson et al., 2016), personality assessments (e.g., Huang et al., 2024), and ability tests (Ramsey & Bowling, 2024). In the current white paper, we review the IER literature and provide practical advice that organizational practitioners and researchers can use to mitigate the effects of IER in their data.

Review of the IER Literature

What Is IER?

IER occurs when respondents complete an assessment without following the instructions, reading the questions, or providing thoughtful responses (Huang et al., 2012). Some inattentive respondents, for instance, may respond while distracted or in a hurry; others may only superficially attend to survey content or ignore it entirely. IER can occur in any data collection effort. IER rates vary across study contexts and sample types, but 5% to 15% of survey responses may exhibit notable IER, and far more may exhibit IER for only a portion of the survey (Meade & Craig, 2012). Even small rates of IER may pose a risk to the accuracy of findings (see Credé, 2010; DeSimone et al., 2018; Huang, Liu, & Bowling, 2015).

Why Does IER Matter?

IER leads to scores that do not accurately reflect the construct of interest. Assessments that contain IER will produce misleading research results. Intuitively, respondents who answer questions in ways that do not accurately reflect their true standing on a construct are providing data of questionable utility to practitioners and researchers. Undesirable effects of these misleading responses can manifest in several ways:

1. IER can affect the mean scores observed within a given dataset. Such effects occur because, as a group, inattentive respondents (or a particular respondent who selects responses at random) generally score closer to a scale's midpoint than do attentive respondents (Credé, 2010; Huang et al., 2024; Huang, Liu, & Bowling, 2015). Consider, for example, workers' responses to a 7-point job satisfaction measure. Because most people are satisfied with their jobs (see Spector, 1997), we would expect that attentive respondents as a group would score well above the scale's midpoint value of 4. On the other hand, we would expect that inattentive respondents as a group would score near the scale's midpoint value. As a result, IER would produce results that underestimate the workers' true job satisfaction levels.
2. IER can affect a scale's reliability and validity (DeSimone et al., 2018). The presence of IER, for example, can lower a scale's observed alpha or omega estimates (Huang et al., 2012) and distort a scale's structure within an exploratory factor analysis. This is especially true for scales containing both positively and negatively keyed items (Schmitt & Stuits, 1985).
3. IER can affect the observed relationship between two scales. Depending on the type and prevalence of IER, research has found that IER can either weaken (McGrath et al., 2010) or inflate (Credé, 2010; Huang et al., 2024; Huang, Liu, & Bowling, 2015) observed relationships between measures. If two assessments both exhibit similar levels of IER, their observed relationship may be artificially inflated. Huang, Liu, and Bowling (2015) provide a detailed discussion of the mechanisms that produce such inflation effects. When IER changes an assessment's reliability, criterion-related validity, or both, stakeholders may make decisions (e.g., selection, promotion) based on untrustworthy results.

How to Assess IER

There are many ways to identify IER. Some respondents will provide the same answer to all questions, so some indices focus on identifying response patterns. Other respondents will select random answers, so some indices focus on response speed or accuracy. Because IER takes multiple forms (see DeSimone et al., 2018; Meade & Craig, 2012), effective detection of IER may require the use of multiple assessment approaches. Although this paper does not provide a full review of IER assessment methods, several comprehensive reviews are available (see Curran, 2016; DeSimone et al., 2015; Meade & Craig, 2012), and more recently, Huang et al. (2025) reviewed over a decade of research on IER detection in organizational science, highlighting empirical inconsistencies and providing recommendations for applied assessments.. In this paper, we review three categories of IER indices: (a) attention check, (b) response time, and (c) response pattern indices. This subset of indices can effectively assess most forms of IER and each can be easily adopted within most organizational assessment efforts.[2] In Tables 1 through 3, we provide additional information about each of these types of indices, including recommendations for best practice.

Table 1
Summary of Attention Check Indices

IER index	Description and example	Advantages	Limitations	Best practice considerations
Infrequency	Infrequency measures consist of special items that have ostensibly correct responses that are uniform across all participants (e.g., "I have never brushed my teeth"; see Meade & Craig, 2012). A researcher infers that IER has occurred whenever a participant provides incorrect responses to multiple infrequency items.	• Research supports the construct validity of infrequency indices as measures of IER (e.g., Huang, Bowling, et al., 2015; Meade & Craig, 2012). • Infrequency measures are easy to implement and they require no specialized statistical skill.	• The infrequency approach requires the inclusion of additional items, thus increasing participant burden. • Assessment stakeholders (e.g., management) may object to infrequency items that include bizarre content.	• Include multiple infrequency items that are distributed evenly throughout the assessment. • Avoid using infrequency items that are overly conspicuous when placed among the assessment's substantive items.
Instructed response	Instructed-response measures consist of special items that direct participants to respond in a particular way (e.g., "To indicate that you are paying attention, leave this item blank"; see Kam & Chan, 2018). A researcher infers that IER has occurred whenever a participant provides incorrect responses to multiple instructed-response items.	• Research supports the construct validity of instructed-response indices as measures of IER (e.g., Kam & Chan, 2018). • Instructed-response measures are easy to implement and they require no specialized statistical skill. • Instructed-response items are easy for researchers to score, because each has one clearly correct response.	• The instructed-response approach requires the inclusion of additional items, thus increasing participant burden. • Assessment stakeholders (e.g., management) may object to instructed-response items, because they cannot be seamlessly embedded into study questionnaires.	• Include multiple instructed-response items that are distributed evenly throughout the assessment.

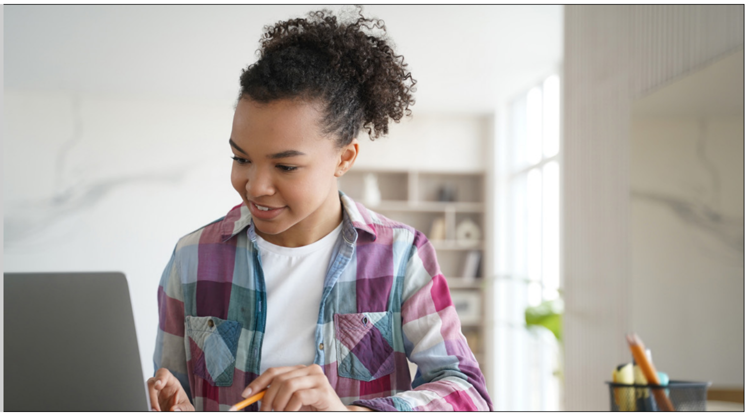
Table 2
Summary of Response Time Indices

IER index	Description and example	Advantages	Limitations	Best practice considerations
Total completion time	Total completion time is the amount of time (e.g., number of seconds) that have elapsed between when a participant begins responding to an assessment and when he or she submits his or her responses (see Bowling et al., 2023).	• Research moderately supports the construct validity of the total completion time index as a measure of IER (see Bowling et al., 2023). • Total completion time is easy to implement via most electronic data collection platforms (e.g., Qualtrics), and it requires no specialized statistical skill. • Total completion time requires no additional items, thus minimizing participant burden. • Assessment stakeholders (e.g., management) are unlikely to object to the total completion time index.	• Total completion time unfortunately, provides a coarse assessment of IER, since it only captures consistently fast responding throughout the entire survey. • Total completion time can only be computed when data are collected using a platform (e.g., Qualtrics) that records completion time.	• Researchers should apply a natural log transformation to address positively skewed total completion time data (see Bowling et al., 2023).
Page time	Page time assesses the extent to which participants respond excessively fast to individual questionnaire pages (see Huang et al., 2012).	• Research supports the construct validity of the page time index as a measure of IER (see Bowling et al., 2023). • Page time is easy to implement via most electronic data collection platforms (e.g., Qualtrics), and it requires no specialized statistical skill. • Page time requires no additional items, thus minimizing participant burden. • Assessment stakeholders (e.g., management) are unlikely to object to the page time index.	• Page time can only be computed when data are collected using a platform (e.g., Qualtrics) that records completion time.	• To compute page time, researchers must administer their assessment across multiple questionnaire pages and record response time values separately for each page.

Table 3
Summary of Response Pattern Indices

IER index	Description and example	Advantages	Limitations	Best practice considerations
Long strings	Long string indices equates IER with the presence of identical responses across several consecutive items (see Johnson, 2005).	• Research moderately supports the construct validity of the long string approach to assessing IER (see Meade & Craig, 2012). • The long strings approach requires no additional items, thus minimizing participant burden. • Assessment stakeholders (e.g., management) are unlikely to object to the long string approach.	• Captures only one specific form of IER—providing *identical* responses across consecutive items. • Long string indices do not capture excessive variability across a participant's responses.	• Long string indices are only appropriate when used to examine responses to consecutive items that reflect two or more distinct constructs.
Intra-individual response variability (IRV) index	The IRV index (see Dunn et al., 2018) is a participant's within-person standard deviation of responses across an item set. It thus equates IER with low variability across a given participant's responses.	• Research supports the construct validity of the IRV index as a measure of IER (see Dunn et al., 2018). • IRV captures forms of excessive response uniformity that are overlooked by long string indices. • The IRV index requires no additional items, thus minimizing participant burden. • The IRV index is easy to implement and requires no specialized statistical skill. • Assessment stakeholders (e.g., management) are unlikely to object to the IRV index.	• IRV does not capture excessive variability across a participant's responses.	• The IRV index is only appropriate when used to examine responses to consecutive items that reflect two or more distinct constructs.

Attention check indices are designed to detect IER using "easy" questions with obvious answers.



1. ***Attention check indices.*** Attention check indices are designed to detect IER using "easy" questions with obvious answers. The rationale is that if participants are paying attention, then they will get these questions right (see Huang, Bowling, et al., 2015; Maniaci & Rogge, 2014; Meade & Craig, 2012). Here, we review two types of attention-check measures: (a) infrequency indices (also known as "bogus items"), and (b) instructed-response indices.

1a. Infrequency items. The term "infrequency" refers to the idea that people who select response options rarely chosen by others may be engaging in IER. Sample infrequency items include "I have never brushed my teeth" (Meade & Craig, 2012) and "I work 14 months in a year (Huang, Bowling, et al., 2015). It is safe to assume that both of these statements are false for all respondents. Therefore, respondents who agree with either of these items are likely not paying attention and therefore exhibiting IER. Responses to infrequency items are coded dichotomously as either "correct" (scored a "0") or "incorrect" (scored a "1"). In the sample items above, indicating any kind of disagreement (from *Slightly Disagree* to *Strongly Disagree*) would be scored as "correct," whereas failing to do so would result in a score of "incorrect" (Huang, Bowling et al., 2015). The researcher computes an overall infrequency score by summing these recoded values. Ideally, a given study questionnaire would include multiple infrequency items to assess IER at various points throughout the assessment, evenly dispersing these items across the assessment. Infrequency items should be written to blend seamlessly with the study's substantive items, and the researcher should include a mix of positively scored and negatively scored infrequency items (in the former, agreement indicates IER; in the latter, disagreement indicates IER).

1b. Instructed-response indices. Instructed-response indices direct participants to provide a particular response (see Kam & Chan, 2018). Example instructed-response items include "For this item, please select slightly agree" and "To show that you are paying attention, please leave this item blank." Instructed-response items are also scored dichotomously: Participants receive a score of "1" for each item to which they do not follow the instructions; they receive a score of "0" for each item to which they follow the instructions. The researcher computes a total instructed-response score by summing the number of incorrect responses. As with infrequency items, researchers should distribute multiple instructed-response items throughout their assessments.

2. ***Response time indices.*** Response time indices simply measure how long a respondent takes when completing an assessment (or portions of an assessment). The rationale is that if participants are "speeding" through the assessment, then they are likely exhibiting IER. This is because reading the questions and responding thoughtfully should require more time than skimming or ignoring the questions. Response time indices are easily implemented using electronic data collection platforms (e.g., Qualtrics) and can be used without the respondents knowing they are being timed. This approach to assessing IER can be used in two different ways: (a) total completion time and (b) page time (see Bowling et al., 2023). Total completion time is the amount of time (e.g., number of seconds) that has elapsed between when a participant begins responding to an assessment and when he or she submits his or her responses. This measure provides a single, coarse assessment of IER that captures consistently fast responding throughout the entire survey. As a result, this index cannot detect respondents who have sped through much of an assessment but lingered on some parts of it (or taken mid-assessment breaks).

Page time, in contrast, requires researchers to record the amount of time participants spend responding to individual questionnaire pages (see Huang et al., 2012). The response time for each page is then recoded to reflect whether the participant was excessively fast in responding. Although page time is still somewhat coarse (i.e., it cannot detect respondents who take midpage breaks), it is more capable of detecting when and where a respondents may have exhibited IER during the course of a survey. A convenient cutoff proposed by Huang et al. (2012) was a rate of 2 seconds per item[3]: on each page, faster response rates are coded "1" to indicate the presence of IER, whereas slower response rates are coded "0" to indicate the absence of IER. When completing a questionnaire page that includes 20 items, for example, participants who spend less than 40 seconds responding would be flagged as having engaged in IER on that page. A total page time score is computed by counting the number of pages on which a participant violated the 2-seconds-per-item rule.

3. Response pattern indices. Response pattern indices measure IER by identifying respondents who consistently respond in very similar ways, even to distinct questions. The rationale is that the "easiest" or "lowest effort" method of completing a survey simply involves selecting the same response to each item or alternating between responses in a pattern. Here, we describe two types of response pattern indices—(a) the long string index and (b) the intraindividual response variability (IRV) index.

3a. The long string index. The long string index simply counts the number of consecutive identical responses a respondent selects (see Johnson, 2005). Selecting *strongly disagree* for 20 consecutive items, for instance, could indicate that a given participant has engaged in IER, particularly when those 20 items (a) contain both positively- and negatively keyed items or (b) assess two or more conceptually distinct constructs. For practitioners familiar with statistical programs and analysis, the longstring index can be automatically computed in R using the "careless" package. The computation is relatively straightforward. Simply count the number of consecutive identical responses on each survey page. Once you have these counts for each page, the longstring index can be computed as either the average of these counts or the maximum of these counts (Meade & Craig, 2012).

3b. The IRV index. The IRV index (Dunn et al., 2018, see also Marjanovic et al., 2015) is calculated for each respondent as the standard deviation of their scores across items on the assessment. This index is relatively simple to compute and interpret for practitioners familiar with basic statistics. Like the longstring index, IRV is intended to capture overly consistent responding. IRV, however, is more flexible than is the longstring index. Consider the prior example of a respondent indicating "strongly disagree" (coded as a score of "1") to 20 consecutive items. This would result in an IRV of zero. If that respondent responded to a single item with "disagree" (coded as a "2"), it would break the longstring, but the IRV would remain low (0.22). If the respondent alternated between "strongly disagree" and "disagree," IRV would still remain somewhat low (0.51). The range of IRV depends on the number of questions and response options. (For a 20-question, 5-option assessment, it ranges from 0 to 2.05.) Dunn et al. (2018) recommend reviewing the IRV scores for every respondent and screening those whose scores are particularly low. However, as with the long string index, the item sets used to compute the IRV index should ideally comprise several distinct constructs and/or both positively and negatively keyed items. This is important because even attentive participants may display little variability in their responses to single-construct assessment with all items scored in the same direction.

IER Mitigation Strategies

In this section, we describe strategies for minimizing the effects of IER. Such strategies have important practical implications, because they may help practitioners and researchers improve the accuracy of the conclusions drawn from their data. These strategies involve (a) the detection and deletion of high-IER data, and (b) the prevention of IER.

Detection and Deletion of High-IER Data

One approach to combatting IER involves assessing IER (e.g., by using one or more of the indices we described in the previous section) and then omitting data from any respondent who has engaged in excessive levels of IER (see Huang et al., 2012). Several practical issues should be considered when using this approach. First, the value of the cut score used to distinguish between attentive and inattentive responses should be determined before data analysis begins. Cut scores, ideally, should be tailored to each individual study. To facilitate this process, it may be necessary to first conduct a pilot study that includes the same substantive measures and uses the same type of sample that will be used in the main study. Ideally, the pilot data would be collected from highly motivated respondents and, when feasible, include a subset of respondents that the researcher has asked to engage in IER (see Huang et al., 2012, Study 1). Such pilot data could be used to identify a specific cut score value that accurately differentiates between attentive and inattentive respondents. When a pilot study is not feasible, a rational approach (see Huang et al., 2012) may be adopted to identify and exclude the more egregious end of the IER continuum, which has been shown to exert disproportional impact on study conclusions (Huang & DeSimone, 2021).

Although effective, the detect-and-delete approach has two main limitations:

1. the approach requires the omission of some respondents

2. it may result in the systematic loss of particular kinds of respondents

Although effective, the detect-and-delete approach has two main limitations. First, the approach requires the omission of some respondents, thus lowering the study's final sample size and threatening statistical power. Users of psychological measures may reasonably be wary of deleting collected data, particularly when a significant amount of time, effort, or resources is required to collect data from each individual participant. Practitioners should always carefully weigh the costs of decreasing sample size against the potential costs of including low-quality data in an analysis. Second, the detect-and-delete approach may result in the systematic loss of particular kinds of respondents, thus harming the representativeness of the study's final sample. Such effects on sample representativeness occur because some participants may be predisposed to habitually engage in IER, particularly those who are low in conscientiousness, extraversion, agreeableness, or emotional stability (Bowling et al., 2016, Study 4), or those high in implicit aggression (DeSimone et al., 2020). Therefore, deleting respondents who exhibit IER may bias assessment results in undetectable ways and/or limit the extent to which assessment results are representative or widely generalizable.

Prevention of IER

A second mitigation approach focuses on IER prevention. Such an approach offers potential advantages over the detect-and-delete approach because prevention minimizes the loss of respondents due to excessive IER. Retaining more respondents yields larger samples, higher statistical power, and potentially less bias in assessment results. The most promising IER prevention strategies introduce incentives for responding attentively. It may be possible to incentivize attentiveness in some settings using warnings (see Huang et al., 2012). Respondents could, for example, be warned at the beginning of the assessment that IER is being monitored and that some form of punishment may be imposed on those who are found to have engaged in IER (Bowling et al., 2021). Stern warnings may not be advisable when assessing employees or customers, especially when participation is voluntary. In these situations, a benign warning may be more appropriate (Huang, Bowling, et al., 2015). In this latter type of warning the researcher simply notes that he or she will monitor the participant's level of attentiveness without threatening any sanctions for engaging in IER. Alternatively, a researcher may use rewards to incentivize attentive responding. A gift card, for example, could be awarded to respondents who are not flagged for IER, or the most attentive respondents could be entered into a raffle for a larger prize (Gibson & Bowling, 2020).

Beyond incentives, two administrative features have demonstrated some potential to prevent IER (see Bowling et al., 2021; Meade & Craig, 2012). First, we recommend administering assessments in person with a proctor or other respondents in the room. Compared with remote (i.e., online) assessments, in-person assessments may yield more attentive responses for two reasons. First, direct interpersonal contact with respondents may allow test administrators to more effectively apply persuasion tactics that encourage attentive responses. Second, within an in-person assessment, administrators have greater control over distractions. They can, for instance, ask respondents to surrender their smartphones during the assessment administration. Controlling distractions is much more difficult within a remote setting.

The use of shorter assessments can also help prevent IER. Simply put, lengthy assessments may cause test takers to become fatigued, thus undermining the motivation and ability needed to provide attentive responses (see Bowling et al., 2021). Distractions may also become more alluring over the course of longer assessments. We do not recommend sacrificing measurement quality for the sake of shortening an assessment, but practitioners should explore strategies to maximize respondent focus. Ideas include breaking the assessment into multiple shorter sessions, finding high-quality short forms of assessments, or scheduling assessments in a manner that ensures respondents are not in a hurry to complete the assessment as fast as possible.

Summary

IER is a widespread problem that compromises data quality. Ignoring IER poses high risk for unknowingly collecting low-quality data, which can lead to flawed conclusions and poor organizational decisions. The current white paper describes a set of IER measures that are effective and easily employed within most data collection efforts. These measures help identify and potentially screen out inattentive participants, thereby improving data quality. Additionally, we suggest designing data collection efforts to mitigate IER. In doing so, organizational researchers and practitioners can be more confident in the quality of the data they collect, the analyses they conduct, the conclusions they draw, and the decisions they make.

Notes

[1] Researchers have used various terms to refer to this behavior, including “insufficient effort responding” (Huang et al., 2012), “careless responding” (Meade & Craig, 2012), “random responding” (Credé, 2010), and “participant inattention” (Maniaci & Rogge, 2014).

[2] We omitted some IER indices from our current discussion because they have questionable validity. Self-reported IER (see Maniaci & Rogge, 2014), for instance, may be an ineffective assessment approach because its validity depends upon the tenuous assumption that inattentive participants will provide careful and honest responses to self-report items, including questions related to their response effort or quality. We omitted other IER indices from our discussion because they may be impractical for many applied settings. Inconsistency indices (e.g., psychometric synonyms and antonyms; see Meade & Craig, 2012), for instance, may be impractical because they require the inclusion of many item pairs—a requirement that is untenable when using brief assessments.

[3] Research has indicated that an appropriate cut off may depend on item length (Wood et al., 2017). Two seconds is a reasonable cut off for most self-report questionnaires with item stems containing a single phrase or sentence. However, a shorter cut off may be appropriate for single-word items such as those on the PANAS (Watson & Clark, 1999). Bowling et al. (2023) found that the page time index provided superior detection of IER than did total completion time. Furthermore, they suggested the use of 1.5-seconds-per-item cutoff for short or medium length items and 2-seconds-per-item cut off for relatively longer items. A longer cut off may be necessary for longer questions such as situational judgment tests or conditional reasoning tests where the item stem contains multiple sentences or paragraphs.

References

Bowling, N. A., Gibson, A. M., Houpt, J. W., & Brower, C. K. (2021). Will the questions ever end? Person-level increases in careless responding during questionnaire completion. *Organizational Research Methods*, *24*, 718-738. https://doi.org/10.1177/1094428120947794

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology, 111*, 218-229. https://doi.org/10.1037/pspp0000085

Bowling, N. A., Huang, J. L., Brower, C. K., & Bragg, C. B. (2023). The quick and the careless: The construct validity of page time as a measure of insufficient effort responding to surveys. *Organizational Research Methods, 26, 323-352.* https://doi.org/10.1177/10944281211056520

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement, 70*, 596-612. https://doi.org/10.1177/0013164410366686

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19. https://doi.org/10.1016/j.jesp.2015.07.006

DeSimone, J. A., Davison, H. K., Schoen, J. L., & Bing, M. N. (2020). Insufficient effort responding as a meaningful construct and a partial function of latent aggression. *Organizational Research Methods, 23*, 154-180. https://doi.org/10.1177/1094428118799486

DeSimone, J. A., DeSimone, A. J., Harms, P. D., & Wood, D. (2018). The differential impacts of two forms of insufficient effort responding. *Applied Psychology: An International Review, 67,* 309-338. https://doi.org/10.1111/apps.12117

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36,* 171-181. https://doi.org/10.1002/job.1962

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, *33*, 105-121. https://doi.org/10.1007/s10869-016-9479-0

Gibson, A. M., & Bowling, N. A. (2020). The effects of questionnaire length and behavioral consequences on careless responding. *European Journal of Psychological Assessment, 36*, 410-420. https://doi.org/10.1027/1015-5759/a000526

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. https://doi.org/10.1037/0021-9010.75.5.581

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology, 30*, 299-311. https://doi.org/10.1007/s10869-014-9357-6

Huang, J. L., Bowling, N. A., McLarty, B. D., Kluemper, D. H., & Wang, Z. (2024). Confounding effects of insufficient effort responding across survey sources: The case of personality predicting performance. *Organizational Research Methods*, *27*, 681-712. https://doi.org/10.1177/10944281231212570

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. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., & DeSimone, J. A. (2021). Insufficient effort responding as a potential confound between survey measures and objective tests. *Journal of Business and Psychology, 36*, 807-828. https://doi.org/10.1007/s10869-020-09707-2

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*, 828-845.http://dx.doi.org/10.1037/a0038510

Huang, J. L., Wang, Z., Huang, R., Wu, D., & Shi, H. (2025). Insufficient effort responding in management research: A critical review and future directions. *Journal of Management.* Advance online publication. https://doi.org/10.1177/01492063251330268

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

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, *39*, 103-129. https://doi.org/10.1016/j.jrp.2004.09.009

Kam, C. C. S., & Chan, G. H. H. (2018). Examination of the validity of instructed response items in identifying careless respondents. *Personality and Individual Differences*, *129*, 83-87. https://doi.org/10.1016/j.paid.2018.03.022

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India, 2*, 49–55. https://doi.org/10.1007/s13171-019-00164-5

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61-83. https://doi.org/10.1016/j.jrp.2013.09.008

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences, 84*, 79-83. https://doi.org/10.1016/j.paid.2014.08.021

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. https://doi.org/10.1037/a0019216

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

Morgeson, F. P., Spitzmuller, M., Garza, A. S., & Campion, M. A. (2016). Pay attention! The liabilities of respondent experience and carelessness when making job analysis judgments. *Journal of Management, 42*, 1904–1933. https://doi.org/10.1177/0149206314522298

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872. https://doi.org/10.1016/j.jesp.2009.03.009

Ramsey, M. C., & Bowling, N. A. (2024). Building a bigger toolbox: The construct validity of existing and proposed measures of careless responding to cognitive ability tests. *Organizational Research Methods*. Advance online publication. https://doi.org/10.1177/10944281231223127

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

Spector, P. E. (1997). *Job satisfaction: Applications, assessment, causes and consequences*. Sage. https://doi.org/10.4135/9781452231549

Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form. Iowa City, IA: University of Iowa. https://doi.org/10.17077/48vt-m4t2

Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science, 8*, 454-464. https://doi.org/10.1177/1948550617703168

Yentes, R. D., & Wilhelm, F. (2018). careless: Procedures for computing indices of careless responding. *R package version*, *1*, 2018. https://cran.r-project.org/web/packages/careless/index.html