Behavior Research Methods (2024) 56:7790–7813
https://doi.org/10.3758/s13428-024-02452-x

ORIGINAL MANUSCRIPT

Validating the IDRIS and IDRIA: Two infrequency/frequency scales for detecting careless and insufficient effort survey responders

Cameron S. Kay[1,2]

Accepted: 16 May 2024 / Published online: 8 July 2024
© The Psychonomic Society, Inc. 2024

Abstract

To detect careless and insufficient effort (C/IE) survey responders, researchers can use infrequency items – items that almost no one agrees with (e.g., "When a friend greets me, I generally try to say nothing back") – and frequency items – items that almost everyone agrees with (e.g., "I try to listen when someone I care about is telling me something"). Here, we provide initial validation for two sets of these items: the 14-item *Invalid Responding Inventory for Statements* (IDRIS) and the 6-item *Invalid Responding Inventory for Adjectives* (IDRIA). Across six studies ($N_1 = 536$; $N_2 = 701$; $N_3 = 500$; $N_4 = 499$; $N_5 = 629$, $N_6 = 562$), we found consistent evidence that the IDRIS is capable of detecting C/IE responding among statement-based scales (e.g., the HEXACO-PI-R) and the IDRIA is capable of detecting C/IE responding among both adjective-based scales (e.g., the Lex-20) and adjective-derived scales (e.g., the BFI-2). These findings were robust across different analytic approaches (e.g., Pearson correlations; Spearman rank-order correlations), different indices of C/IE responding (e.g., person-total correlations; semantic synonyms; horizontal cursor variability), and different sample types (e.g., US undergraduate students; Nigerian survey panel participants). Taken together, these results provide promising evidence for the utility of the IDRIS and IDRIA in detecting C/IE responding.

Keywords Infrequency/frequency items · Scale validation · Careless/insufficient effort responding · Random responding · Data quality

Introduction

If you have ever conducted a survey, you have probably encountered so-called "careless and insufficient effort" (C/IE) responders (see Curran, 2016). C/IE responders are participants who provide responses to items on a survey that are unrelated to the content of those items. For example, a participant who does not read any items on a survey and instead selects the same response option for every item would be classified as a C/IE responder, as would a participant who responds according to some predetermined pattern or who selects responses at random.

The purpose of the present set of six studies is to validate two scales for detecting C/IE responders: the 14-item *Invalid Responding Inventory for Statements* (IDRIS; Kay, 2021) and the 6-item *Invalid Responding Inventory for Adjectives* (IDRIA; Kay, 2023). Despite recently being included in the *Comprehensive Infrequency/Frequency Item Repository* (CIFR; Kay & Saucier, 2023) – an online database of 660 infrequency/frequency items – these two scales have not undergone any form of formal validation. Here, we provide this validation by examining the IDRIS and IDRIA in relation to several well-established indices of C/IE responding.

Background

Although estimates of the prevalence of C/IE responding vary widely (e.g., Berry et al., 1992; Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012), many researchers place the number somewhere around 10% (see Curran, 2016). This would not be a problem if C/IE responding had little to no impact on data quality, but researchers have consistently found that including even small numbers of C/IE responders in one's data can have dramatic consequences. For example, Woods (2006) found that data with as little as 10% C/IE responders can introduce additional factors

✉ Cameron S. Kay
cameronstuartkay@gmail.com

[1] Psychology Department, Union College, 807 Union Street, Schenectady, NY 12308, USA

[2] Department of Psychology, University of Oregon, 1227 University of Oregon, Eugene, OR 97403, USA

in otherwise unidimensional data (see also Arias et al., 2020; DeSimone et al., 2018; Schmitt & Stults, 1985). Other researchers have, likewise, found that including C/IE responders in one's data can artificially inflate (Cornell et al., 2012; DeSimone et al., 2018; Holtzman & Donnellan, 2017; Zorowitz et al., 2023) and artificially deflate (Credé, 2010; Hough et al., 1990; Huang et al., 2015a; Oppenheimer et al., 2009) observed effect sizes, leading to increased type I and type II error rates, respectively.

Given the threat that C/IE responders pose to data quality, researchers have, unsurprisingly, devoted considerable time and effort to developing methods for detecting these participants (see Curran, 2016; Ward & Meade, 2023). Researchers have, for example, developed ways of detecting these responders by looking at the length of time it takes participants to complete surveys (Huang et al., 2012), the variability participants show in their responses to surveys (Thalmayer & Saucier, 2014), the number of items in a row that participants provide the same response to on surveys (Johnson, 2005), and the movement of participants' cursors when completing surveys (Pokropek et al., 2023).

One additional method that has recently been receiving greater attention is the so-called infrequency/frequency-item method (see Curran, 2016; Ward & Meade, 2023). The idea behind this method is simple: participants who are engaged in C/IE responding will be more likely to agree with items that most people disagree with and disagree with items that most people agree with. Following this method, researchers include in their surveys items that are known to be endorsed relatively infrequently (e.g., "I will be punished for meeting the requirements of my job"; Huang et al., 2015b) and items that are known to be endorsed relatively frequently (e.g., "It feels good to be appreciated"; Maniaci & Rogge, 2014). They then screen their data for participants who show a consistent pattern of agreeing and disagreeing with the infrequency and frequency items, respectively.[1,2]

One of the first instantiations of the infrequency/frequency-item method was Washburne's (1935; see also Hartshorne & May, 1928) *objectivity* scale. As part of his social adjustment measure for children, Washburne included items that asked respondents whether they had engaged in behaviours that almost everyone has or has not engaged in. For example, participants were asked whether they had ever broken or lost something that belonged to someone else and whether they were always on time for school and other appointments. These items were, according to Washburne, meant to assess a respondent's ability to accurately report their conduct and feelings, with the ultimate goal being to guard against "intentional and unintentional inaccuracies in the answering of the questionnaire" (p. 126).

Eight years after the introduction of Washburne's objectivity scale, Hathaway and McKinley (1943) introduced what would become one of the most widely used infrequency/frequency scales, even today. In the first edition of the *Minnesota Multiphasic Personality Inventory* (MMPI), Hathaway and McKinley included an "F scale", which comprised 44 infrequency items (e.g., "Evil spirits possess me at times") and 20 frequency items (e.g., "I get angry sometimes"). The infrequency items were those that were endorsed by fewer than 10% of visitors to the hospital and outpatient department at the University of Minnesota. The frequency items were those that were endorsed by greater than 90% of the visitors to the hospital and outpatient department. The scale was originally intended to detect C/IE responders, participants who had misinterpreted items, and response sheets that had been miscoded. Later, it would be recognized that the scale could also be used to detect participants who were trying to exaggerate their symptoms (i.e., "faking bad"; Meehl & Hathaway, 1946). Since its debut, the MMPI has undergone several revisions (e.g., the MMPI-3; Ben-Porath & Tellegen, 2020) and has, directly or indirectly, influenced the creation of numerous infrequency/frequency measures.

One set of measures indirectly inspired by the MMPI is the IDRIS (Kay, 2021) and the IDRIA (Kay, 2023). Despite having similar names, the IDRIS and IDRIA are intended for two distinct use cases. The IDRIS was developed to be used with scales composed of statements, such as the *HEXACO-PI-R* (e.g., "In social situations, I'm usually the one who makes the first move"; Ashton & Lee, 2005; Lee & Ashton, 2004). Accordingly, the IDRIS includes seven infrequency statements (e.g., "I am older than my parents") and seven frequency statements (e.g., "I can remember the names of most of my close family members") (Appendix 1). The IDRIA, on the other hand, was developed to be used with scales composed of adjectives, such as the *Big Five*

[1] See Kay and Saucier (2023) for a discussion of the use of the phrase "frequency item" versus "negatively-keyed infrequency item".

[2] Of course, there may be some participants who genuinely agree with the statement "I will be punished for meeting the requirements of my job" and some participants who genuinely disagree with the statement "It feels good to be appreciated". People invariably differ, and it is nearly impossible to find items that all participants will agree or disagree with. In fact, Hathaway and McKinley (1943) raised this issue in the publication manual for the first edition of the *Minnesota Multiphasic Personality Inventory* (MMPI), writing that there are people who "may admit to disliking children and not believing their mother was a good woman" (p. 8). Moreover, participants often interpret items in overly liberal ways. For example, Curran and Hauser (2019) found that some participants agree with the item "I have been to every country in the world" because they had been to many countries. However, as more infrequency/frequency items are included in a survey, the likelihood that a participant will be incorrectly flagged as a C/IE responder becomes increasingly less likely. As a case in point, only 0.01% of people would agree with four infrequency items that each have a false positive rate of 10.00%, provided the items are not correlated with each other.

Mini-Markers (e.g., "Bold"; Saucier, 1994). Accordingly, the IDRIA includes three infrequency adjectives (e.g., "triangular") and three frequency adjectives (e.g., "mortal") (Appendix 2).

Although the IDRIS and IDRIA have not been formally validated, both scales have a number of features that make them promising as infrequency/frequency scales. To start, the infrequency items and frequency items that make up the IDRIS and IDRIA are relatively infrequent and frequent, respectively. For infrequency/frequency scales to work, most people (or, more specifically, most non-C/IE responders) have to disagree with the infrequency items and agree with the frequency items (see Hathaway & McKinley, 1943). This appears to be the case for the IDRIS and IDRIA. Specifically, Kay and Saucier (2023) demonstrated that, on a five-point Likert scale, participants provided an average response of 1.08 to 1.42 to the infrequency items from the IDRIS and an average response of 4.44 to 4.72 to the frequency items from the IDRIS. Likewise, participants provided an average response of 1.47 to 1.91 to the infrequency items from the IDRIA and an average response of 4.60 to 4.81 to the frequency items from the IDRIA.[3]

A second desirable feature of the IDRIS and IDRIA is that both scales are nonproprietary.[4] The scales can, therefore, be accessed by researchers who may not otherwise have the funds to purchase measurement manuals or scale booklets, a common barrier to conducting research for early-career researchers and researchers from countries without established funding agencies. Being nonproprietary also means the IDRIS and IDRIA are free to be reworded, rearranged, and otherwise modified by researchers, saving the scales from the psychometric purgatory that often befalls proprietary measures (see Goldberg et al., 2006).

A third desirable feature of the IDRIS and IDRIA is that both scales include equal numbers of infrequency items and frequency items. Most extant infrequency/frequency scales include more infrequency items than frequency items (e.g., Hathaway & McKinley, 1943), with some scales being composed entirely of infrequency items (e.g., Beach, 1989; Huang et al., 2015b). In some ways, including more infrequency items than frequency items makes sense. Like most participants (Cronbach, 1946), C/IE responders are more likely to agree with items than disagree with items (Johnson, 2005). Since infrequency items are designed to detect improbable *agreement*, it is understandable that researchers would feel compelled to include more infrequency items than frequency items in their scales. However, including only infrequency items make these scales unable to detect C/IE responders who, for whatever reason, tend to disagree with items. Based on estimates from Johnson (2005), this could be as much as 30.71% of C/IE responders.

A fourth desirable feature of the IDRIS and IDRIA is that both scales were developed with subtlety in mind. Ideally, infrequency/frequency items should be as subtle as possible (see Curran, 2016), as this minimizes the chance that a C/IE responder's attention will be drawn to the items long enough for them to be recognized as attention check items. As noted by Kay and Saucier (2023), many extant infrequency/frequency items include conspicuous linguistic features that undermine their subtlety. These conspicuous linguistic features include (a) proper nouns (e.g., "I own *Starbucks*"; Dunn et al., 2018), (b) uncommon words (e.g., "I am paid biweekly by *leprechauns*"; Meade & Craig, 2012), (c) numbers (e.g., "I can run *2* miles in *2* minutes"; Huang et al., 2015b), and (d) unusual punctuation (e.g., "I lie 100 % of the time"; Dunn et al., 2018). The items from the IDRIS were developed specifically to avoid these features.

A fifth desirable feature relates only to the IDRIA, and it is that the IDRIA can be used among adjective-based scales. Although adjective-based scales have fallen somewhat out of favor over the last several decades, many are still in widespread use. For example, the Big Five Mini-Markers (Saucier, 1994) – a set of 40 adjectives for assessing the Big Five personality traits – was cited (if not necessarily used) 140 times in 2020 alone. This is not to mention the Midlife Development Inventory (Lachman & Weaver, 1997), which has been used in numerous large-scale data collection efforts over the past three decades (e.g., Beals et al., 2003; Juster & Suzman, 1995; Ryff et al., 2018). The recent development of a number of adjective-based scales (e.g., the Narcissistic Grandiosity Scale; Crowe et al., 2016) also suggests the field may be going through something of an adjective-based-scale renaissance. As far as we know, the IDRIA (Kay, 2023) is the only infrequency/frequency scale that is composed entirely of adjectives and is, therefore, the only infrequency/frequency scale that is appropriate for use among adjective-based scales.

[3] These estimates were obtained using the statement version of the IDRIA, which is discussed in further detail toward the end of the introduction.

[4] Admittedly, most modern infrequency/frequency scales are also nonproprietary. Nevertheless, we believe it is important to highlight this feature of the IDRIS and IDRIA here because many older (and widely-used) infrequency/frequency scales remain proprietary. Take the *Deviant Responding* subscale from the *Psychopathic Personality Inventory – Revised* (PPI-R; Lilienfeld & Widows, 2005) – which the IDRIS is closely modelled after – as an example. To use the Deviant Responding subscale from the PPI-R, researchers would need to purchase the measurement manual for $135.00 and either buy scale booklets ($108.00 for a pack of 25) or pay an online administration fee ($0.35 per administration with a minimum fee of $350.00). This is despite the content of the items being left largely unchanged since 1996.

With only a small amount of modification, the IDRIA is also the only infrequency/frequency scale that is appropriate for use among adjective-derived scales. By "adjective-derived scales", we mean statement-based scales that were originally developed using adjectives. Take the Big Five Inventory (John & Srivastava, 1999) as an example. The scale is ostensibly statement-based, but, from looking at the items, it is evident that they were developed, at least in part, from adjectives. The item "I see myself as someone who is original, comes up with new ideas" can, for instance, be boiled down to the adjective "original". By reversing this process, the IDRIA items can be made to work with adjective-derived scales (see Appendix 2). The IDRIA item "asleep" can, for example, be changed to "I am asleep, not awake", which would fit in quite well with the other items from the Big Five Inventory. In contrast, the IDRIS item "If I heard a loud noise behind me, I would turn around to see what it was" would be quite conspicuous if administered alongside the items from the Big Five Inventory.

The current study

Despite their promising features, it is yet unclear whether the IDRIS and IDRIA are able to actually detect C/IE responding. This is, of course, a problem: detecting C/IE responding is the whole purpose of these scales. To remedy this issue, we conducted six studies. Study 1 evaluated the validity of the IDRIS by examining its association with a number of previously validated indices of C/IE responding. Study 2 further evaluated the validity of the IDRIS, while also providing an initial validation of the IDRIA. Study 3 served as an additional replication of the results for the IDRIS from Study 1 and Study 2, while Study 4 served as an additional replication of the results for the IDRIA from Study 2. In Study 5, we moved beyond the samples of American undergraduate students used in the prior four studies to examine whether the IDRIS and IDRIA are capable of detecting C/IE responding in broad samples of adults from the US, India, and Nigeria. In Study 6, we used a sample from Amazon's Mechanical Turk (MTurk) to investigate whether the IDRIS and IDRIA are able to detect C/IE survey responding on a popular on-demand data collection platform. Moreover, we examined whether the IDRIS and IDRIA are associated with C/IE responding to the same degree as a previously validated infrequency/frequency scale. In five of the six studies, we also fit receiver operating characteristic (ROC) curves to identify optimal cut thresholds for identifying C/IE respondents using the IDRIS and IDRIA.

Study 1

Study 1 was intended to provide a preliminary investigation of the validity of the IDRIS. To that end, we examined the associations of the IDRIS with several common indices of C/IE responding, including response durations, long strings of identical responses, intra-individual response variabilities, person-total correlations, responses to psychometric synonyms, and responses to psychometric antonyms. We also tested whether the IDRIS is able to predict whether a given participant will provide fake e-mail addresses when asked to provide contact information for three informants.

Method

Participants and procedures

Five hundred undergraduate students (70.00% women; 27.60% men; *M* age = 19.52; *SD* age = 2.40) completed the IDRIS as part of a larger survey administered at the University of Oregon. This sample size was selected to fit the needs of a separate project. Nevertheless, a sample of this size would have a 99.98% probability of detecting a large effect ($r = .30$; Funder & Ozer, 2019) with a two-tailed alpha level of .001 when such an effect existed. The survey included 264 statements spread across four blocks. The first block included 77 items, the second block included 82 items, the third block included 76 items, and the fourth block included 29 items. These items were drawn from a diverse set of measures, including the 40-item *Narcissistic Personality Inventory* (Raskin & Hall, 1979), the 32-item *Uniqueness Scale* (Snyder & Fromkin, 1977), and the 20-item *Desirability of Control Scale* (Burger & Cooper, 1979)[5].

Materials

IDRIS The IDRIS ($\bar{r}_{ij} = .30$; $\alpha = .85$) items were intermixed with the other items in the survey. Specifically, two of the IDRIS items were included in the first block of the survey and four of the IDRIS items were included in the second, third, and fourth blocks of the survey. Each block included equal numbers of infrequency items and frequency items. In order to create an index of C/IE responding, the frequency items were reverse-scored and averaged together with the infrequency items. Higher scores on the resulting composite indicated a greater likelihood of C/IE responding. Participants responded to the IDRIS, as well as the filler items, on a five-point Likert scale (– 2 = "Strongly disagree"; 2 = "Strongly agree").

[5] This data was previously reported by Kay (2021) and Kay and Slovic (2023).

Response duration Response duration refers to the length of time it takes a participant to respond to a survey (Bowling et al., 2021; Huang et al., 2012; Wise & Kong, 2005), with shorter durations being indicative of C/IE responding. In the present study, some participants had extremely long response durations (e.g., 85 min), potentially due to leaving the survey open while they completed other tasks. In order to address this issue, we recoded the response times for the 10% slowest responders as missing values (see Meade & Craig, 2012).

Long-string index The long-string index captures the longest string of identical responses provided by each participant (Johnson, 2005). In this case, longer strings are indicative of C/IE responding. In the present study, we used the *longstring* function from the *careless* package (Yentes & Wilhelm, 2021) to produce a long-string index across the entire survey, as well as within each of the four blocks.

Intra-individual response variability Intra-individual response variability (IRV) refers to the standard deviation of a participant's responses to a set of items (Thalmayer & Saucier, 2014; see also Dunn et al., 2018). Low IRV values are indicative of C/IE responding.[6] In the present study, we used the *irv* function from the *careless* package (Yentes & Wilhelm, 2021) to calculate the IRV for each participant. As with the long-string index, we calculated IRV scores across the entire survey, as well as within each of the four blocks.

Person-total correlation A person-total correlation is a correlation between a given participant's responses and the average participant's responses (Donlon & Fischer, 1968; see also Curran, 2016). Although participants should not all exhibit the same pattern of responses, they should, if they are responding validly, provide responses that are at least somewhat similar to the responses of others. As such, low person-total correlations are indicative of C/IE responding. We used the *profile* function from the *panoply* package (Kay, 2019) to calculate person-total correlations for each participant in the present survey. As with the long-string index and IRV scores, we produced person-total correlations across the entire survey, as well as within each of the four blocks.

[6] Some researchers argue that C/IE responders will show *greater* variability than non-C/IE responders because they tend to select randomly from across the entire range of response options (Marjanovic et al., 2015). In the present study, as in prior studies (Dunn et al., 2018), IRV scores were negatively associated with the other indices of C/IE responding, indicating that lower (not higher) IRV scores are associated with C/IE responding.

Psychometric-synonyms and psychometric-antonyms indices The psychometric synonyms index represents the within-person correlation between highly positively correlated pairs of items, while the psychometric antonyms index represents the within-person correlation between highly negatively correlated pairs of items (Meade & Craig, 2012). Accordingly, smaller positive correlations on the psychometric synonyms index are indicative of C/IE responding, while smaller negative correlations on the psychometric antonyms index are indicative of C/IE responding. In the present study, the psychometric synonyms included any pairs of items that demonstrated a correlation more extreme than .60. The psychometric antonyms were initially intended to include any pairs of items that demonstrated a correlation more extreme than – .60 (see Meade & Craig, 2012). This threshold was decreased to – .50, however, as there were not enough pairs of items to produce the psychometric antonyms index when the threshold was – .60. The psychometric synonyms index was based on 33 pairs of items and, after reducing the threshold, the psychometric antonyms index was based on 3 pairs of items. The indices were calculated using the *psychsyn* function from the *careless* package (Yentes & Wilhelm, 2021).

Fake informant e-mail addresses After completing the four blocks of the survey, the participants were asked to provide e-mail addresses for three people who knew them well enough to accurately rate their personalities. The first author reviewed the e-mail addresses and flagged any participants that provided e-mail addresses that (a) included the participants' first or last names, suggesting the participants were recommending themselves, (b) were Gmail e-mail addresses and less than six characters long (i.e., the minimum length for Gmail e-mail addresses), (c) indicated that the participant was purposefully not cooperating with the instructions (e.g., providing the e-mail addresses "no@email.com", "nope@email.com", and "nothanks@email.com"), (d) were all identical (e.g., providing the e-mail address "jsmith@email.com" for all three informants), (e) included the names of celebrities (e.g., providing the e-mail addresses "markhamill@email.com", "carriefisher@email.com", and "harrisonford@email.com"), (f) included the names of fictional characters (e.g., providing the e-mail addresses "drmanhattan@email.com", "rorschach@email.com", and "ozymandias@email.com"), (g) followed a pattern (e.g., providing the e-mail addresses "aaaaaa@email.com", "bbbbbb@email.com", and "cccccc@email.com"), or (h) were otherwise improbable (e.g., providing the e-mail addresses "htnkjl@email.com", "mnhjktl@email.com", and "bchpkljh@email.com").

Table 1 Zero-order correlations of the IDRIS with the indices of C/IE responding (Study 1)

	IDRIS				
	All	Block 1	Block 2	Block 3	Block 4
Duration	– .40*	– .25*	– .32*	– .34*	– .36*
Long string	.48*	.35*	.37*	.39*	.42*
Block 1	.44*	.39*	.35*	.35*	.36*
Block 2	.45*	.36*	.38*	.36*	.36*
Block 3	.49*	.30*	.36*	.43*	.44*
Block 4	.45*	.29*	.28*	.34*	.50*
IRV	– .55*	– .39*	– .42*	– .48*	– .48*
Block 1	– .50*	– .42*	– .41*	– .39*	– .40*
Block 2	– .56*	– .39*	– .43*	– .48*	– .48*
Block 3	– .55*	– .34*	– .40*	– .49*	– .49*
Block 4	– .43*	– .30*	– .26*	– .36*	– .44*
Person-total correlation	– .75*	– .47*	– .62*	– .64*	– .63*
Block 1	– .57*	– .44*	– .48*	– .43*	– .46*
Block 2	– .67*	– .41*	– .59*	– .57*	– .51*
Block 3	– .71*	– .30*	– .53*	– .64*	– .63*
Block 4	– .50*	– .18*	– .31*	– .45*	– .52*
Psychometric synonyms	– .57*	– .30*	– .44*	– .51*	– .49*
Psychometric antonyms	.23*	.05	.21*	.23*	.18*

* $p < .001$. IRV = Intra-individual response variability

Results and discussion

Descriptive statistics for all of the Study 1 variables are provided in the Supplementary Material, as are the zero-order correlations among the variables.

Associations of the IDRIS with the C/IE indices

Consistent with our expectations, the higher a person scored on the IDRIS, the more likely they were to (a) speed through the survey ($r = – .40$, $p < .001$), (b) provide long strings of identical responses ($r = .48$, $p < .001$), (c) exhibit low response variabilities ($r = – .55$, $p < .001$), (d) depart from the average pattern of responses ($r = – .75$, $p < .001$), (e) respond differently to psychometrically synonymous items ($r = – .57$, $p < .001$), and (f) respond similarly to psychometrically antonymous items ($r = .23$, $p < .001$) (Table 1)[7]. When asked to provide e-mail addresses for the three informants, participants scoring high on the IDRIS were also more likely to provide fake e-mail addresses, $\chi^2(1, N = 500) = 16.22$, $p < .001$. Specifically, the odds of the participant providing a fake e-mail increased by 2.05 (95% CI [1.45, 2.91]) times for every one-unit higher they scored on the IDRIS ($b = 0.72$, 95% CI [0.37, 1.07], $SE = 0.18$, Wald = 4.02, $p < .001$). Overall, these results indicate that the IDRIS is a useful predictor of C/IE responding.

Optimal cut-off thresholds for identifying C/IE responders using the IDRIS

To identify the best cut-off scores for identifying C/IE responders using the IDRIS, we fit an ROC curve (see Stanislaw & Todorov, 1999; Table 2). An ROC curve illustrates the trade-off between sensitivity – the proportion of correctly classified positive cases among all positive cases – and specificity – the proportion of correctly classified negative cases among all negative cases – for all levels of some classifier. The area under the curve (AUC) provides an index of how well the scale is able to differentiate positive from negative cases. In the present study, positive and negative cases of C/IE responding were identified by subjecting the six continuous indices of C/IE responding to a K-means clustering algorithm.[8] The greatest average Silhouette score (i.e., how close a participant was to their own cluster relative to other clusters) was achieved for a two-cluster solution (.46). According to thresholds reported by Mandrekar (2010), the IDRIS had an excellent classification ability (AUC = .97). The cut-off score with the best sensitivity and specificity for the IDRIS was – .96.

Study 2

The purpose of Study 2 was two-fold. First, we wanted to replicate and extend the findings from Study 1. For the most part, we did this by examining the association of the IDRIS with the same indices of C/IE responding used in Study 1. However, instead of having participants recommend three informants, we had participants complete a self-report measure of C/IE responding. Second, we wanted to provide an initial evaluation of the validity of the IDRIA. We did this by examining the IDRIA in

[7] Given a number of the variables were highly skewed, we reanalyzed the present data using both (1) Spearman rank-order correlations and (2) Pearson correlations after implementing exponential- and log-based transformations. The results of these reanalyses are provided in the Supplementary Material. The conclusions that can be drawn from the reanalyses are identical to the conclusions that can be drawn from the analyses reported here.

[8] In each study where K-means clustering was performed, the indices of C/IE responding were preprocessed in two ways. First, they were standardized. Second, missing values were imputed using the *missRanger* function from the *missRanger* package (Mayer, 2021) with a maximum of 10 chaining iterations, 10,000 trees, unlimited tree depth, and 3 variables randomly sampled at each split.

Table 2 Receiver operating characteristic curve results predicting the K-means clusters and self-report measure of C/IE responding from the IDRIS and IDRIA

		K-means clusters						Self-report measure of C/IE responding					
	Scale	Threshold	Adjusted	AUC	Sensitivity	Specificity	Accuracy	Threshold	Adjusted	AUC	Sensitivity	Specificity	Accuracy
IDRIS													
Study 1	5-point	– 0.96	– 0.96	.97	.93	.92	.92	-	-	-	-	-	-
Study 2	5-point	– 0.93	– 0.93	.99	.97	.91	.91	– 0.79	– 0.79	.89	.86	.90	.90
Study 3	5-point	– 0.96	– 0.96	.97	.95	.97	.97	-	-	-	-	-	-
Study 6	7-point	– 0.43	– 0.29	.83	.90	.70	.84	– 0.21	– 0.14	.73	.83	.54	.63
IDRIA													
Study 2	9-point	– 0.75	– 0.38	.88	.77	.84	.83	– 0.75	– 0.38	.80	.71	.79	.79
Study 4	5-point	– 0.75	– 0.75	.91	.84	.91	.90	-	-	-	-	-	-
Study 6	7-point	– 0.58	– 0.38	.86	.83	.79	.82	– 0.42	– 0.28	.74	.84	.55	.65

AUC = area under the ROC curve; adjusted = threshold scaled to be on a 5-point response scale

relation to the same indices used to further test the validity of the IDRIS.

Method

Participants and procedures

Seven hundred one undergraduate students (67.76% women; 28.10% men; *M* age = 19.48; *SD* age = 1.92) completed the IDRIS and IDRIA as part of a larger survey administered at the same university as in Study 1. Again, this sample size was selected to fit the needs of a separate project. A sample of this size would have a 99.99% probability of detecting a large effect (r = .30; Funder & Ozer, 2019) with a two-tailed alpha level of .001 when such an effect existed. The survey included 377 statements spread across three blocks: the first block included 125 statements, the second block included 124 statements, and the third block included 128 statements. As in Study 1, the statements were pulled from a variety of measures, including the 100-item *HEXACO-PI-R* (Ashton & Lee, 2005; Lee & Ashton, 2004), the 64-item *Self-Report Psychopathy Scale - 4* (Paulhus et al., 2016), and the 20-item *Mach-IV* (Christie & Geis, 1970). The survey also included 101 adjectives, which were presented in a fourth block. The adjectives included the 6 items from the IDRIA and the 95 items from the *Lexical Factor Model of Personality - 20* (Lex-20; Saucier & Iurino, 2020).

Materials

IDRIS See Study 1 for a full description of the IDRIS ($\bar{r}_{ij}$ = .31; α = .84). In the present study, four of the IDRIS items were included in each of the first, second, and third blocks of the survey. As in Study 1, each block included equal numbers of infrequency items and frequency items.

Due to a coding error, the first IDRIS item was not collected for the first 500 respondents. We used the *missRanger* function from the *missRanger* package (Mayer, 2021) to impute the missing values using a chained random forest model with a maximum of 10 chaining iterations, 10,000 trees, unlimited tree depth, and 3 variables randomly sampled at each split. The conclusions that can be drawn from the analyses were the same if we used the imputed values or if we based the IDRIS scores on the average of the 13 IDRIS items that were administered to these participants.

IDRIA The IDRIA ($\bar{r}_{ij}$ = .19; α = .56) items were intermixed randomly with the adjectives from the Lex-20 in the fourth block. As with the IDRIS, the frequency items were reverse scored and averaged together with the infrequency items to produce an index of C/IE responding. Participants responded to these adjectives, as well as the Lex-20, using a nine-point response scale (– 4 = "extremely inaccurate"; 4 = "extremely accurate").

Response duration See Study 1 for a full description of how response duration was assessed. In Study 2, we added a separate timer to each block, allowing us to calculate each participant's overall and block-specific response durations.

Long-string index See Study 1 for a full description of how the long-string index was calculated. We produced a long-string index across the three statement-based survey blocks, as well as within each of the four blocks.

Intra-individual response variability See Study 1 for a full description of how IRV was calculated. We calculated IRVs across the three statement-based survey blocks, as well as within each of the four blocks.

Person-total correlation See Study 1 for a full description of how the person-total correlations were calculated. We produced person-total correlations across the three statement-based survey blocks, as well as within each of the four blocks.

Psychometric-synonyms and psychometric-antonyms indices See Study 1 for a full description of how psychometric synonyms and psychometric antonyms indices were calculated. In the present study, we calculated separate psychometric synonyms and psychometric antonyms indices for the statements (Block 1, Block 2, and Block 3) and for the adjectives (Block 4). A correlation of .60 was sufficient for calculating the psychometric synonyms index for the statements and the adjectives. A correlation of – .55 was required to generate the psychometric antonyms for the statements and the adjectives. In the end, the psychometric synonyms index for the statements was based on 42 pairs of items, and the psychometric synonyms index for the adjectives was based on 18 pairs of items. The psychometric antonyms index for the statements was based on 3 pairs of items and the psychometric antonyms index for the adjectives was based on 6 pairs of items.

Self-report measure of C/IE responding Participants responded to a single item at the end of the survey assessing their self-reported levels of C/IE responding. The item was similar to that used by Aust and colleagues (2012) and Meade and Craig (2012). Specifically, the participants were asked, "Is there any reason we should exclude your responses from our analyses (e.g., you did not respond to the survey questions truthfully; you selected answers at random)?" Participants could respond by either selecting "Yes – my survey SHOULD be thrown out" or "No – my survey SHOULD NOT be thrown out." We decided to include this item only after we had already collected data from 75 participants.

Results and discussion

Descriptive statistics for all of the Study 2 variables are provided in the Supplementary Material, as are the zero-order correlations among the variables.

Associations of the IDRIS and IDRIA with the C/IE indices

As in Study 1, participants scoring high on the IDRIS were more likely to (a) speed through the survey ($r = -.27$, $p < .001$), (b) provide long strings of identical responses ($r = .50$, $p < .001$), (c) exhibit low response variabilities ($r = -.68$, $p < .001$), (d) depart from the average pattern of responses ($r = -.68$, $p < .001$), (e) respond differently to psychometrically synonymous items ($r = -.65$, $p < .001$), and (f) respond similarly to psychometrically antonymous items ($r = .18$, $p < .001$) (Table 3; Fig. 1)[9]. Participants high on the IDRIS were also more likely to indicate that their data should be discarded, $\chi^2(1, N = 629) = 94.85$, $p < .001$. Specifically, for every one-unit higher a participant scored on the IDRIS, the odds that they said their data should be excluded increased by 20.68 (95% CI [10.42,

Table 3 Zero-order correlations of the IDRIS and IDRIA with the indices of C/IE responding (Study 2)

	IDRIS				IDRIA
	Block 1/2/3	Block 1	Block 2	Block 3	Block 4
Duration	– .27*	– .20*	– .29*	– .21*	– .13
Block 1	– .33*	– .33*	– .31*	– .24*	– .17*
Block 2	– .37*	– .31*	– .36*	– .29*	– .18*
Block 3	– .48*	– .30*	– .42*	– .47*	– .20*
Block 4	– .45*	– .29*	– .40*	– .42*	– .24*
Long string	.50*	.38*	.47*	.43*	.29*
Block 1	.42*	.42*	.39*	.30*	.22*
Block 2	.49*	.37*	.49*	.39*	.30*
Block 3	.51*	.34*	.46*	.47*	.28*
Block 4	.47*	.34*	.45*	.41*	.29*
IRV	– .68*	– .47*	– .67*	– .59*	– .45*
Block 1	– .65*	– .56*	– .60*	– .52*	– .39*
Block 2	– .65*	– .42*	– .68*	– .55*	– .46*
Block 3	– .67*	– .40*	– .63*	– .63*	– .43*
Block 4	– .58*	– .38*	– .59*	– .50*	– .51*
Person-total correlation	– .68*	– .58*	– .58*	– .57*	– .33*
Block 1	– .59*	– .55*	– .49*	– .48*	– .24*
Block 2	– .63*	– .50*	– .57*	– .50*	– .33*
Block 3	– .67*	– .45*	– .54*	– .63*	– .33*
Block 4	– .56*	– .44*	– .51*	– .46*	– .36*
Psychometric synonyms					
Block 1/2/3	– .65*	– .47*	– .55*	– .58*	– .31*
Block 4	– .58*	– .38*	– .49*	– .54*	– .36*
Psychometric antonyms					
Block 1/2/3	.18*	.08	.15*	.18*	.13*
Block 4	.37*	.23*	.35*	.34*	.26*

* $p < .001$. *IRV* = Intra-individual response variability

[9] As in Study 1, the conclusions remained the same if the data was reanalyzed using (1) Spearman rank-order correlations and (2) Pearson correlations after implementing exponential- and log-based transformations (see the Supplementary Material).

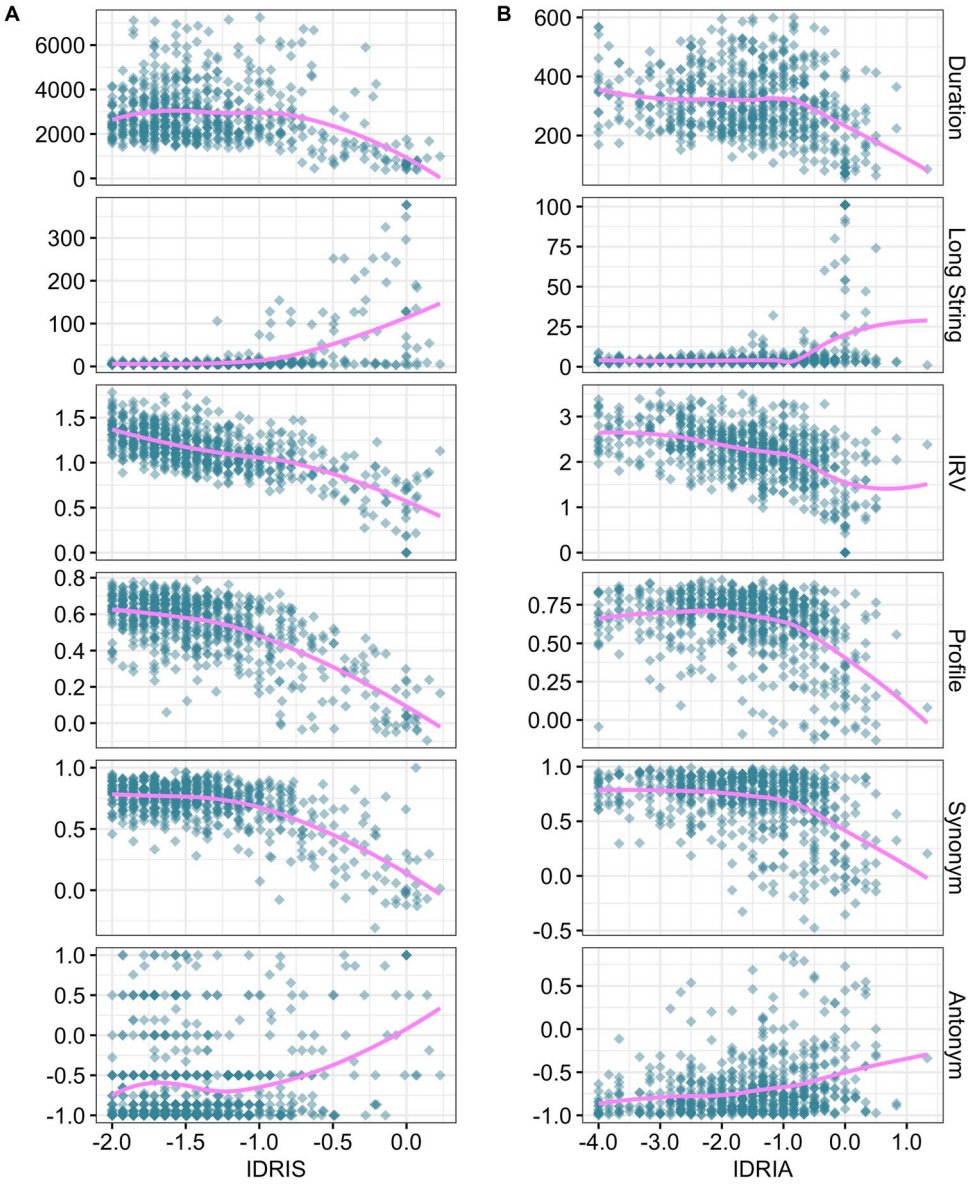

Fig. 1 Scatterplots depicting the association of the **A** IDRIS and **B** IDRIA scores with response durations, the long string index, intra-individual response variabilities, person-total correlations, psychometric synonyms, and psychometric antonyms in Study 2

45.09]) times (b = 3.03, 95% CI [2.34, 3.81], SE = 0.37, Wald = 8.16, $p < .001$) (Fig. 2).

Participants high on the IDRIA were also more likely to (a) speed through the survey ($r = -.24$, $p = .001$), (b) provide long strings of identical responses ($r = .29$, $p < .001$), (c) exhibit low response variabilities ($r = -.51$, $p < .001$), (d) depart from the average pattern of responses ($r = -.36$, $p < .001$), (e) respond differently to psychometrically synonymous items ($r = -.36$, $p < .001$), and (f) respond similarly to psychometrically antonymous items ($r = .26$, $p = .001$) (Table 3; Fig. 1). They were also more likely to indicate that their data should be discarded, $\chi^2(1, N = 629) = 39.62$, $p < .001$. For every one-unit higher a participant scored on the IDRIA, the odds that they said their data should be excluded increased by 3.69 (95% CI [2.36, 6.03]) times (b = 1.30, 95% CI [0.86, 1.80], SE = 0.24, Wald = 5.47, $p < .001$) (Fig. 2).

Overall, these findings indicate that, as in Study 1, the IDRIS is a valid predictor of C/IE responding. It also provides preliminary evidence that the IDRIA is a valid predictor of C/IE responding.

Optimal cut-off thresholds for identifying C/IE responders using the IDRIS and IDRIA

To identify the best cut-off scores for identifying C/IE responders using the IDRIS and IDRIA, we fit ROC curves

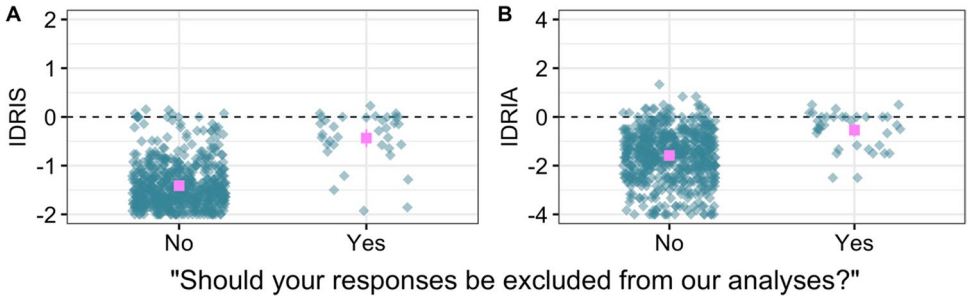

Fig. 2 Plots depicting the average **A** IDRIS and **B** IDRIA scores among participants who indicated that their data should and should not be excluded. The *error bars* represent 95% confidence intervals

(Table 2). As in Study 1, positive and negative cases of C/IE responding were identified by subjecting the continuous indices of C/IE responding to a K-means clustering algorithm. For the IDRIS, we used the combined C/IE responding indices from Block 1, 2, and 3, and, for the IDRIA, we used the C/IE responding indices from Block 4. In both cases, the greatest average Silhouette score was achieved for a two-cluster solution ($Silhouette_{1/2/3} = .56$; $Silhouette_4 = .55$). The IDRIS had an outstanding classification ability (AUC = .99) and the IDRIA had an excellent classification ability (AUC = .88). The optimal cut-off score was – .93 for the IDRIS and – .75 for the IDRIA.

We also reran the ROC curve analysis using the self-report measure of C/IE responding as the outcome variable. The results were largely the same as when we used K-means clustering. Both the IDRIS (AUC = .89) and IDRIA (AUC = .80) had excellent classification abilities. The optimal cut-off score was – .79 for the IDRIS and – .75 for the IDRIA.

Study 3

The purpose of Study 3 was to provide further validation of the IDRIS. The indices used to validate the IDRIS in Study 3 are the same as those used in Study 2, but we did not assess self-reported C/IE responding. Instead, we calculated a semantic antonyms index.

Method

Participants and procedures

Five hundred thirty-six undergraduate students (64.55% women; 31.53% men; *M* age = 19.80; *SD* age = 2.35) completed the IDRIS as part of a larger survey administered at the same university as in Study 1 and Study 2. As in the prior studies, the sample size was selected to fit the needs of a separate project. A sample of this size would have a 99.99% probability of detecting a large effect ($r = .30$; Funder & Ozer, 2019) with a two-tailed alpha level of .001 when such an effect existed. The survey included a single block of 88 statements, including 14 items from the IDRIS, 60 items from the *HEXACO-PI-R* (Ashton & Lee, 2005; Lee & Ashton, 2004), and 14 items from a Semantic Antonyms Set.

Materials

See Study 1 for a full description of the IDRIS ($\bar{r}_{ij} = .42$; $\alpha = .90$). In the present study, we made several changes to the IDRIS items. Most of these changes were intended to either increase clarity or make the infrequency items more infrequent and the frequency items more frequent. As an example of increasing clarity, we updated the item, "I often say goodbye before I end a phone call" to "I often say *some form* of goodbye *right* before I end a phone call". We wanted to make it clear that participants should endorse this item if they use any words of farewell before hanging up the phone (e.g., "see you later"), not just the literal word "goodbye". As an example of making the infrequency items more infrequent and the frequency items more frequent, we updated the item "It should be illegal to intentionally kill another person" to "It should be illegal to intentionally kill an innocent person". We were concerned that some people would disagree with this item solely because they support capital punishment. A side-by-side comparison of the items used in Study 1, Study 2, Study 3, and Study 5 can be found in the Supplementary Material.

Response duration See Study 1 for a full description of how response duration was assessed.

Long-string index See Study 1 for a full description of how the long-string index was calculated.

Intra-individual response variability See Study 1 for a full description of how IRV was calculated.

Person-total correlation See Study 1 for a full description of how the person-total correlations were calculated.

Psychometric-synonyms and psychometric-antonyms indices See Study 1 for a full description of how the psychometric synonyms and psychometric antonyms indices were calculated. In the present study, a correlation of .50 was necessary to generate the psychometric synonyms index. A correlation of – .60 was sufficient for generating the psychometric antonyms index. In the end, the psychometric synonyms index was based on four pairs of items, and the psychometric antonyms index was based on three pairs of items.

Semantic antonyms The idea behind a semantic antonyms index is the same as a psychometric antonyms index but, instead of selecting pairs of items based on observed correlations in one's data, researchers include items in their surveys that are judged to be semantically antonymous a priori (see Goldberg & Kilkowski, 1985). Here, we included seven pairs of semantically antonymous items (e.g., "I go through money quickly" and "I am good at saving money"), hereafter referred to as the *Semantic Antonyms Set*. These items turned out to be both semantically antonymous *and* psychometrically antonymous: the correlations between the pairs of items in the Semantic Antonyms Set ranged from – .31 to – .71. A full list of the items in the Semantic Antonyms Set, as well as their correlations, can be found in the Supplementary Material.

Results and discussion

Descriptive statistics for all of the Study 3 variables are provided in the Supplementary Material, as are the zero-order correlations among the variables.

Associations of the IDRIS with the C/IE indices

Consistent with our expectations, the higher a person scored on the IDRIS, the more likely they were to (a) speed through the survey ($r = -.58$, $p < .001$), (b) provide long strings of identical responses ($r = .52$, $p < .001$), (c) exhibit low response variabilities ($r = -.65$, $p < .001$), (d) depart from the average pattern of responses ($r = -.62$, $p < .001$), (e) respond differently to psychometrically synonymous items ($r = -.34$, $p < .001$), (f) respond similarly to psychometrically antonymous items ($r = .32$, $p < .001$), and (g) respond similarly to semantically antonymous items ($r = .41$, $p < .001$) (Table 4)[10]. As in Study 1 and Study 2, the present results indicate that the IDRIS is a useful predictor of C/IE responding.

[10] As in Study 1 and Study 2, the conclusions remained the same if the data was reanalyzed using (1) Spearman rank-order correlations and (2) Pearson correlations after implementing exponential- and log-based transformations (see the Supplementary Material).

Table 4 Zero-order correlations of the IDRIS with the indices of C/IE responding (Study 3)

	IDRIS
Duration	– .58*
Long string	.52*
IRV	– .65*
Person-Total Correlation	– .62*
Psychometric Synonyms	– .34*
Psychometric Antonyms	.32*
Semantic Antonyms	.41*

* $p < .001$. *IRV* = Intra-individual response variability

Optimal cut-off thresholds for identifying C/IE responders using the IDRIS

As in the prior two studies, we fit an ROC curve to identify the best cut-off score for identifying C/IE responders using the IDRIS (Table 2). Positive and negative cases of C/IE responding were, again, identified by subjecting the continuous indices of C/IE responding to a K-means clustering algorithm. In this case, the greatest average Silhouette score was achieved for a two-cluster solution (Silhouette = .51). The IDRIS had an outstanding classification ability (AUC = .97), and the optimal cut-off score was – .96.

Study 4

The purpose of Study 4 was to provide further validation of the IDRIA. However, instead of administering the IDRIA items in their adjectival form, we presented them in a statement-based form. The purpose of this change was to examine whether the IDRIA would be able to detect C/IE responding among adjective-*derived* scales (e.g., the Big Five Inventory - 2). We examined the IDRIA in relation to the same indices used in Study 3, as well as two instructed response items (e.g., "Select disagree for this statement").

Method

Participants and procedures

Four hundred ninety-nine undergraduate students (63.13% women; 33.87% men; *M* age = 19.63; *SD* age = 2.05) completed the IDRIA as part of a larger survey administered at the same university as in the first three studies. As in the prior studies, the sample size was selected to fit the needs of a separate project. A sample of this size would have a 99.99% probability of detecting a large effect ($r = .30$; Funder & Ozer, 2019) with a two-tailed alpha level of .001 when such an effect existed. After responding to several blocks of items relevant to a separate project, participants were administered a

single block of 52 statements, including 6 statements adapted from the IDRIA, 30 statements from the *Big Five Inventory – 2 Short Form* (Soto & John, 2017b), 14 statements from the Semantic Antonyms Set, and 2 instructed response statements.

Materials

IDRIA See Study 2 for a full description of the IDRIA ($\bar{r}_{ij} = 25$; $\alpha = 63$). In order to administer the IDRIA adjectives as statements, we added text that was intended to clarify (but, critically, not change) the underlying meaning of the items (Appendix 2). For example, "asleep" became "I am asleep, not awake". Participants responded to this version of the IDRIA on a five-point Likert scale (– 2 = "Strongly disagree"; 2 = "Strongly agree").

Response duration See Study 1 for a full description of how response duration was assessed.

Long-string index See Study 1 for a full description of how the long-string index was calculated.

Intra-individual response variability See Study 1 for a full description of how IRV was calculated.

Person-total correlation See Study 1 for a full description of how the person-total correlations were calculated.

Psychometric-synonyms and psychometric-antonyms indices See Study 1 for a full description of how the psychometric synonyms and psychometric antonyms indices were calculated. In the present study, a correlation of .55 was necessary to generate the psychometric synonyms and a correlation of – .55 was necessary to generate the psychometric antonyms. In the end, the psychometric synonyms index was based on four pairs of items, and the psychometric antonyms index was based on four pairs of items.

Semantic antonyms See Study 3 for a full description of how the semantic antonyms index was calculated.

Instructed response items Instructed response items direct participants to select a specific response option from a response scale (see Curran, 2016). Not selecting the requested response option is taken as evidence of C/IE responding. Here, we included two of these items: (1) "Select disagree for this statement" and (2) "Choose the middle response option for this statement".

Results and discussion

Descriptive statistics for all of the Study 4 variables are provided in the Supplementary Material, as are the zero-order correlations among the variables.

Table 5 Zero-order correlations of the IDRIA with the indices of C/IE responding (Study 4)

	IDRIA
Duration	– .33*
Long string	.42*
IRV	– .51*
Person-Total Correlation	– .42*
Psychometric Synonyms	– .26*
Psychometric Antonyms	.18*
Semantic Antonyms	.27*

* $p < .001$. *IRV* = Intra-individual response variability

Associations of the IDRIA with the C/IE indices

Consistent with our expectations, the higher a person scored on the IDRIA, the more likely they were to (a) speed through the survey ($r = – .33$, $p < .001$), (b) provide long strings of identical responses ($r = .42$, $p < .001$), (c) exhibit low response variabilities ($r = – .51$, $p < .001$), (d) depart from the average pattern of responses ($r = – .42$, $p < .001$), (e) respond differently to psychometrically synonymous items ($r = – .26$, $p < .001$), (f) respond similarly to psychometrically antonymous items ($r = .18$, $p < .001$), and (g) respond similarly to semantically antonymous items ($r = .27$, $p < .001$) (Table 5).[11] Moreover, when instructed to respond "disagree" to an item, participants scoring high on the IDRIA were more likely to select something other than "disagree", $\chi^2(1, N = 499) = 68.70$, $p < .001$. Specifically, the odds of the participant responding something other than "disagree" increased by 5.48 (95% CI [3.59, 8.56]) times for every one-unit higher they scored on the IDRIA ($b = 1.70$, 95% CI [1.28, 2.15], $SE = 0.22$, Wald = 7.69, $p < .001$). Likewise, when instructed to select the middle response option to an item, participants scoring high on the IDRIA were more likely to select something other than the middle response option, $\chi^2(1, N = 499) = 65.52$, $p < .001$. Specifically, the odds of the participant selecting something other than the middle response option increased by 16.13 (95% CI [7.53, 39.42]) times for every one-unit higher they scored on the IDRIA ($b = 2.78$, 95% CI [2.02, 3.67], $SE = 0.42$,

[11] As in the prior studies, the conclusions remained the same if the data was reanalyzed using (1) Spearman rank-order correlations and (2) Pearson correlations after implementing exponential- and log-based transformations (see the Supplementary Material).

Wald = 2.78, $p < .001$).[12] As in Study 2, the present results indicate that the IDRIA is a useful predictor of C/IE responding.

Optimal cut-off thresholds for identifying C/IE responders using the IDRIA

As before, we fit an ROC curve to identify the best cut-off score for identifying C/IE responders using the IDRIA (Table 2). Positive and negative cases of C/IE responding were again identified by subjecting the continuous indices of C/IE responding to a K-means clustering algorithm, with the greatest average Silhouette score being achieved by a two-cluster solution (Silhouette = .49). The IDRIA had an outstanding classification ability (AUC = .91). The optimal cut-off score was – .75.

Study 5

The purpose of Study 5 was to examine whether the IDRIS and IDRIA are capable of detecting C/IE responders in broad samples of adults recruited from the US, India, and Nigeria. It is well known that the frequencies of beliefs and behaviors vary across countries (Henrich et al., 2010). If an infrequency/frequency scale is intended to be used in a country other than the one it was developed in, it is, therefore, crucial to evaluate whether the scale is actually able to detect C/IE responding in that country. To our knowledge, this is the first study to compare the validity of non-proprietary infrequency/frequency scales across countries (but see Cheung et al., 1991).

Before proceeding, it should be noted that the data for Study 5 was prescreened by Qualtrics Panels based on two indicators: response durations and long strings of identical responses. Because of this, we have not reported the correlations of the IDRIS and IDRIA with these indices in the main text, as they would not be an accurate reflection of the true associations among these variables. We have, nevertheless, provided these correlations in the Supplementary Material. For the same reason, we did not produce ROC curves in this study.

Method

Participants and procedures

We used Qualtrics Panels to recruit participants from the US (N = 209; 49.76% women; M age = 40.14; SD age = 13.74); India (N = 210; 50.00% women; M age = 35.84; SD age = 11.27); and Nigeria (N = 210; 50.00% women; M age = 34.90; SD age = 10.18). As in the prior studies, the sample sizes were selected to fit the needs of a separate project. Samples of these sizes would have an 87.53% to 87.75% probability of detecting a large effect (r = .30; Funder & Ozer, 2019) with a two-tailed alpha level of .001 when such an effect existed. All participants completed a survey with two blocks. The first block comprised 76 statements, including the IDRIS, the 20-item *Mach-IV* (Christie & Geis, 1970), the 13-item *Narcissistic Personality Inventory - 13* (Gentile et al., 2013), and the 29-item *Self-Report Psychopathy Scale - 4 (Short Form)* (Paulhus et al., 2016). The second block comprised 101 adjectives, including the IDRIA and the Lex-20 (Saucier & Iurino, 2020).

Materials

IDRIS See Study 1 for a full description of the IDRIS (US: $\bar{r}_{ij} = .29, \alpha = .83$; India: $\bar{r}_{ij} = .17, \alpha = .69$; Nigeria: $\bar{r}_{ij} = .14, \alpha = .64$). Due to a discrepancy between our estimate of the duration of the survey and Qualtrics Panels' estimate of the duration of the survey, we administered a reduced set of eight items to the first 25 participants from the US and India. Once we confirmed our estimate was accurate, we administered the full set of 14 items. As in Study 2, we used the *missRanger* function from the *missRanger* package (Mayer, 2021) to impute the missing values using a chained random forest model with a maximum of 10 chaining iterations, 10,000 trees, unlimited tree depth, and 3 variables randomly sampled at each split. The conclusions that can be drawn from the analyses were the same regardless of whether we used the imputed values or based the IDRIS scores on the average of the eight IDRIS items that were administered to these participants.

IDRIA See Study 2 for a full description of the IDRIA (US: $\bar{r}_{ij} = .17, \alpha = .56$; India: $\bar{r}_{ij} = .12, \alpha = .42$; Nigeria: $\bar{r}_{ij} = .13, \alpha = .47$).

Intra-individual response variability See Study 1 for a full description of how IRV was calculated.

[12] A greater proportion of participants failed the "select disagree for this statement" instructed response item (16.83%) than the "choose the middle response option for this statement" instructed response item (4.81%), $\chi^2(1, N = 499) = 79.64$, $p < .001$, $\varphi = .40$. One potential reason for this is that some participants may have interpreted "select disagree for this statement" as an instruction to select either "disagree" or "strongly disagree". When "disagree" or "strongly disagree" are accepted as valid answers, the proportion of participants failing this item decreased (8.82%) and the IDRIA scores were better able to predict which participants would fail the item, $\chi^2(1, N = 499) = 109.11$, $p < .001$. The odds of the participant responding something other than "disagree" or "strongly disagree" increased by 18.58 (95% CI [9.81, 38.17]) times for every one-unit higher they scored on the IDRIA (b = 2.92, 95% CI [2.28, 3.64], SE = 0.34, Wald = 8.47, $p < .001$).

Table 6 Zero-order correlations of the IDRIS and IDRIA with the indices of C/IE responding (Study 5)

	IDRIS (Block 1)			IDRIA (Block 2)		
	US	India	Nigeria	US	India	Nigeria
IRV						
Block 1	– .41*	– .60*	– .49*	– .26*	– .37*	– .17
Block 2	– .44*	– .44*	– .17	– .51*	– .58*	– .40*
Person-total correlation						
Block 1	– .59*	– .54*	– .46*	– .24*	– .23*	– .10
Block 2	– .49*	– .50*	– .14	– .43*	– .53*	– .49*
Psychometric synonyms						
Block 1	– .32*	– .50*	– .31*	– .18	– .35*	– .04
Block 2	– .60*	– .41*	– .19	– .55*	– .44*	– .25*
Psychometric antonyms						
Block 1	.56*	.50*	.16	.30*	.27*	.03
Block 2	.52*	.32*	.20	.48*	.28*	.23*

* $p < .001$. IRV = Intra-individual response variability

Person-total correlations See Study 1 for a full description of how the person-total correlations were calculated.

Psychometric-synonym and psychometric-antonym indices See Study 1 for a full description of how the psychometric synonyms and psychometric antonyms indices were calculated. In the present study, a correlation of .50 was required to generate psychometric synonyms for the statements in each country, and a correlation of .60 was required to generate psychometric synonyms for the adjectives in each country. A correlation of – .25 was required to generate psychometric antonyms for the statements in each country, and a correlation of – .35 was required to generate psychometric antonyms for the adjectives in each country. In the end, the psychometric synonyms scores for the statements were based on 134, 26, and 7 pairs of items for the US, Indian, and Nigerian samples, respectively. The psychometric synonyms scores for the adjectives were based on 53, 41, and 95 pairs of items for the US, Indian, and Nigerian samples, respectively. The psychometric antonyms scores for the statements were based on 17, 5, and 6 pairs of items for the US, Indian, and Nigerian samples, respectively. The psychometric antonyms scores for the adjectives were based on 44, 8, and 185 pairs of items for the US, Indian, and Nigerian samples, respectively.

Results and discussion

Descriptive statistics for all of the Study 5 variables are provided in the Supplementary Material, as are the zero-order correlations among the variables.

Associations of the IDRIS and IDRIA with the C/IE indices

With one exception, participants scoring high on the IDRIS in each country were more likely to (a) exhibit low response variabilities (US: $r = – .41$, $p < .001$; India: $r = – .60$, $p < .001$; Nigeria: $r = – .49$, $p < .001$), (b) depart from the average pattern of responses (US: $r = – .59$, $p < .001$; India: $r = – .54$, $p < .001$; Nigeria: $r = – .46$, $p < .001$), (c) respond differently to synonymous items (US: $r = – .32$, $p < .001$; India: $r = – .50$, $p < .001$; Nigeria: $r = – .31$, $p < .001$), and (d) respond similarly to antonymous items (US: $r = .56$, $p < .001$; India: $r = .50$, $p < .001$; Nigeria: $r = .16$, $p = .023$) (Table 6)[13]. The one exception was that people from Nigeria who scored high on the IDRIS were no more likely to respond similarly to antonymous items, at least at the more conservative alpha level of .001.

Similar to the IDRIS, participants scoring high on the IDRIA were more likely to (a) exhibit low response variabilities (US: $r = – .51$, $p < .001$; India: $r = – .58$, $p < .001$; Nigeria: $r = – .40$, $p < .001$), (b) depart from the average pattern of responses (US: $r = – .43$, $p < .001$; India: $r = – .53$, $p < .001$; Nigeria: $r = – .49$, $p < .001$), (c) respond differently to synonymous items (US: $r = – .55$, $p < .001$; India: $r = – .44$, $p < .001$; Nigeria: $r = – .25$, $p < .001$), and (d) respond similarly to antonymous items (US: $r = .48$, $p < .001$; India: $r = .28$, $p < .001$; Nigeria: $r = .23$, $p = .001$).

[13] As in the prior studies, the conclusions remained the same if the data was reanalyzed using (1) Spearman rank-order correlations and (2) Pearson correlations after implementing exponential- and log-based transformations (see the Supplementary Material).

Taken together, these results indicate that – even when using a dataset that has been prescreened for low response durations and long strings of identical responses – the IDRIS and IDRIA can be used to successfully detect C/IE responding in samples drawn from the US, India, and Nigeria.

Study 6

Study 6 had two purposes. First, we wanted to further replicate and extend the findings from the prior studies. We did this by examining the associations of the IDRIS and IDRIA with the same indices of C/IE responding used in Study 4. However, we used MTurk – a data-collection platform known for its low-quality data (Douglas et al., 2023; Moss & Litman, 2018) – to recruit participants. We also implemented cursor tracking to test whether the IDRIS and IDRIA could detect patterns indicative of C/IE responding in cursor movements. The second purpose of the proposed study was to compare the performance of the IDRIS and IDRIA against an extant infrequency/frequency item measure: the *Deviant Responding* subscale of the *Psychopathic Personality Inventory* (PPI-DR; Lilienfeld, 1994).

We had six preregistered hypotheses (https://osf.io/dpckj/?view_only=680406f9cfdb41a3b12006241973def0).[14] We expected (H1) the IDRIS, IDRIA, and PPI-DR to be positively correlated with each other; (H2) the IDRIS and IDRIA to be negatively correlated with response durations, intra-individual response variabilities, the standard deviations of the participants' horizontal cursor movements, person-total correlations, and the psychometric synonyms index; (H3) the IDRIS and IDRIA to be positively correlated with long strings of identical responses, the psychometric antonyms index, and the semantic antonyms index; (H4) the correlations of the IDRIS and IDRIA with the continuous indices of C/IE responding to not differ significantly from those seen for the PPI-DR; and (H5) higher scores on the IDRIS and IDRIA to be associated with greater odds of failing instructed response items. Critically, we (H6) did not expect higher scores on the IDRIS and IDRIA to be associated with greater odds of reporting that one's data should be excluded via the self-report measure of C/IE responding. This differs from the finding from Study 2, which showed higher scores on the IDRIS and IDRIA were associated with greater odds of reporting that one's data should be excluded. The reason we expected this departure from the prior results was that participants in Study 6 were being paid to participate in the study and may have, consequently, believed that reporting that their data should be excluded would result in their payments being withheld. Participants in Study 2 were undergraduate students taking part in the study for course credit, so this would not have been a concern.

Method

Participants and procedures

Five hundred sixty-two MTurk workers (31.85% women; 66.55% men; *M* age = 32.52; *SD* age = 8.32) completed the IDRIS, IDRIA, and PPI-DR. They were paid at a rate equivalent to $7.25 an hour for their participation. The sample size in this case was based on two power analyses. The first power analysis was for a Pearson *r* correlation. It indicated that 486 participants would be required to detect a .18 correlation – the smallest focal correlation from Study 3 – 80% of the time that such an effect existed with a two-tailed alpha level of .0016. An alpha level of .0016 was used to account for type I error rate inflation resulting from testing 30 associations (i.e., the associations of the IDRIS, IDRIA, and PPI-DR with the 10 indices of C/IE responding). The second power analysis was for the comparison of two dependent correlations with a shared variable, since one of the goals of Study 6 was to compare the associations of the IDRIS and IDRIA with the indices of C/IE responding to the associations of the PPI-DR with the indices of C/IE responding. The power analysis indicated that 304 participants would be required to detect a difference between a correlation of .50 and a correlation of .30 – the smallest difference deemed to be of practical interest – 80% of the time that such an effect existed with a two-tailed alpha level of .0016. In this case, the alpha level of .0016 was used to account for type I error rate inflation resulting from conducting 30 comparisons (i.e., the IDRIS associations versus the IDRIA associations; the IDRIS associations versus the PPI-DR associations; and the IDRIA associations versus the PPI-DR associations). Taking into

[14] We also expected (H2) the PPI-DR to be negatively correlated with response durations, intra-individual response variabilities, the standard deviations of the participants' horizontal cursor movements, person-total correlations, and the psychometric synonyms index; (H3) the PPI-DR to be positively correlated with long strings of identical responses, the psychometric antonyms index, and the semantic antonyms index; (H5) higher scores on the PPI-DR to be associated with greater odds of failing instructed response items; and (H6) higher scores on the PPI-DR to not be associated with greater odds of reporting that one's data should be excluded via the self-report measure of C/IE responding. Given that these hypotheses are not directly relevant to validating the IDRIS and IDRIA, we have reported the results corresponding to these hypotheses in the Supplementary Material.

account both of these power analyses, we opted to collect at least 500 participants.[15]

The survey included 105 statements administered in a single block. The 105 items included 14 items from the IDRIS, 6 items from the statement version of the IDRIA, 10 items from the PPI-DR, 6 items from the Semantic Antonyms Set, 2 instructed response items, 12 items from *agreeableness* subscale of the *International Personality Item Pool's NEO-60* (Maples-Keller et al., 2019), 12 items from the *agreeableness* subscale of the *Big Five Inventory – 2* (Soto & John, 2017a), and 43 items from the *antagonism* subscale of the *Personality Inventory for DSM-5* (Krueger et al., 2012). Participants responded to the statements using a seven-point Likert scale (– 3 = "Strongly disagree"; 3 = "Strongly agree").

Materials

IDRIS See Study 1 for a full description of the IDRIS ($\bar{r}_{ij} = .09$; $\alpha = .58$).

IDRIA See Study 2 and Study 4 for a full description of the IDRIA ($\bar{r}_{ij} = .06$; $\alpha = .28$).

PPI-DR The PPI-DR ($\bar{r}_{ij} = .07$; $\alpha = .48$) includes five infrequency items (e.g., "I occasionally forget my name") and five frequency items (e.g., "I smile at a funny joke at least once in a while"). As with the IDRIS and IDRIA, the frequency items were reverse-scored and averaged together with the infrequency items to create an index of C/IE responding. Higher scores on the resulting composite indicated a greater likelihood of C/IE responding.

Response duration See Study 1 for a full description of how response duration was assessed.

Long-string index See Study 1 for a full description of how the long-string index was calculated.

Intra-individual response variability See Study 1 for a full description of how IRV was calculated.

Horizontal cursor variability The horizontal cursor variability index was based on research from Pokropek and colleagues (2023). However, instead of calculating the sum of each participant's horizontal cursor travel, we calculated the standard deviation of each participant's horizontal cursor travel.[16] Consequently, the index represented how far each participant's cursor typically was from the average position of their cursor in the x-dimension. As with intra-individual response variability, low horizontal cursor variability should be indicative of C/IE responding, as it indicates that participants are only using a relatively small portion of the response scale. Using code provided by Walters (2015), we assessed the position of the participants' cursors every 200 ms in Qualtrics.

Person-total correlation See Study 1 for a full description of how the person-total correlations were calculated.

Psychometric-synonyms and psychometric-antonyms indices See Study 1 for a full description of how the psychometric synonyms and psychometric antonyms indices were calculated. A correlation of .60 was sufficient for calculating the psychometric synonyms index. A correlation of .10 was required to generate the psychometric antonyms.[17] In the end, the psychometric synonyms index was based on 334 pairs of items and the psychometric antonyms index was based on 7 pairs of items.

Semantic antonyms See Study 3 for a full description of how the semantic antonyms index was calculated. To keep administration costs down, only the three best-performing pairs of items from the Semantic Antonyms Set were used in the present study.

Instructed response items See Study 4 for a full description of instructed response items. In the present study, the two instructed response items were: (1) "Choose strongly disagree for this item" and (2) "Select the middle response option for this item".

Self-report measure of C/IE responding See Study 2 for a full description of the self-report measure of C/IE responding.

[15] Study 6 was funded using a faculty research grant. As noted in our preregistration, we intended to continue collecting data until all of the funds in the grant had been depleted. This resulted in a final sample size that was slightly greater than the target of 500 participants.

[16] The IDRIS ($r = .36$, $p < .001$), IDRIA ($r = .34$, $p < .001$), and PPI-DR ($r = .39$, $p < .001$) were also highly associated with the sum of the horizontal distance travelled.

[17] This is not a typo. Only seven pairs of items in the dataset had correlations less than positive .10. Even the three pairs of items from the Semantic Antonyms Set, which were purposefully selected because they assess opposing content, were positively correlated. Specifically, "I am talkative" was positively correlated with "I don't tend to talk a lot" ($r = .32$, $p < .001$); "I go through money quickly" was positively correlated with "I am good at saving my money" ($r = .31$, $p < .001$); and "I do not sleep well" was positively correlated with "I sleep soundly" ($r = .54$, $p < .001$). These findings add to the chorus of research that has raised concerns about the quality of data collected using MTurk (Douglas et al., 2023; Moss & Litman, 2018).

Table 7 Zero-order correlations of the IDRIS, IDRIA, and PPI-DR with the indices of C/IE responding (Study 6)

	IDRIS	IDRIA	PPI-DR
IDRIS	-		
IDRIA	.74*	-	
PPI-DR	.84*	.71*	-
Duration	– .32*$_a$	– .34*$_a$	– .28*$_a$
Long string	.10$_a$	.06$_a$	.10$_a$
IRV	– .61*$_a$	– .57*$_a$	– .59*$_a$
Horizontal cursor variability	– .28*$_a$	– .25*$_a$	– .27*$_a$
Person-total correlation	– .60*$_a$	– .63*$_a$	– .61*$_a$
Psychometric synonyms	– .54*$_a$	– .51*$_a$	– .54*$_a$
Psychometric antonyms	.50*$_a$	.52*$_a$	.52*$_a$
Semantic antonyms	.19*$_a$	.12$_a$	.19*$_a$

* $p < .0016$. Correlations with different subscripts in the same row are significantly different at $p < .0016$ (Hittner et al., 2003). *IRV* = Intra-individual response variability

Results and discussion

Descriptive statistics for all of the Study 6 variables are provided in the Supplementary Material, as are the zero-order correlations among the variables.

Associations of the IDRIS and IDRIA with the C/IE indices

The results were largely consistent with our hypotheses. Consistent with Hypothesis 1, the IDRIS ($r = .84$, $p < .001$) and IDRIA ($r = .71$, $p < .001$) were both highly positively associated with the PPI-DR (Table 7). Consistent with Hypothesis 2 and Hypotheses 3, those scoring high on the IDRIS were more likely to (a) speed through the survey ($r = – .32$, $p < .001$), (b) exhibit low response variabilities ($r = – .61$, $p < .001$), (c) exhibit low horizontal cursor variabilities ($r = – .28$, $p < .001$), (d) depart from the average pattern of responses ($r = – .60$, $p < .001$), (e) respond differently to psychometrically synonymous items ($r = – .54$, $p < .001$), (f) respond similarly to psychometrically antonymous items ($r = .50$, $p < .001$), and (g) respond similarly to semantically antonymous items ($r = .19$, $p < .001$)[18]. Moreover, those scoring high on the IDRIA were more likely to (a) speed through the survey ($r = – .34$, $p < .001$), (b) exhibit low response variabilities ($r = – .57$, $p < .001$), (c) exhibit low horizontal cursor variabilities ($r = – .25$, $p < .001$), (d) depart from the average pattern of responses ($r = – .63$, $p < .001$), (e) respond differently to psychometrically synonymous items ($r = – .51$, $p < .001$), and (f) respond similarly to psychometrically antonymous items ($r = .53$, $p < .001$). Consistent with Hypothesis 4, the associations of the IDRIS and IDRIA with the indices of C/IE responding were not significantly different than those seen for the PPI-DR.

Turning to Hypothesis 5, participants scoring high on the IDRIS were more likely to select something other than "strongly disagree" when instructed to select "strongly disagree" $\left(\chi^2(1, N = 562)\right) = 60.85$, $p < .001$) and something other than the middle response option when instructed to select the middle response option ($\chi^2(1, N = 562) = 75.97$, $p < .001$). Specifically, the odds of a participant selecting something other than "strongly disagree" in the first case increased by 5.36 (95% CI [3.18, 9.73]) times for every one-unit higher they scored on the IDRIS ($b = 1.68$, 95% CI [1.16, 2.28], $SE = 0.29$, Wald = 5.89, $p < .001$). The odds of a participant selecting something other than the middle response option in the second case increased by 5.60 (95% CI [3.45, 9.69]) times for every one-unit higher they scored on the IDRIS ($b = 1.72$, 95% CI [1.24, 2.27], $SE = 0.26$, Wald = 6.53, $p < .001$). Participants scoring high on the IDRIA were also more likely to select something other than "strongly disagree" when instructed to select "strongly disagree" $\left(\chi^2(1, N = 562)\right) = 84.35$, $p < .001$) and something other than the middle response option when instructed to select the middle response option $\left(\chi^2(1, N = 562)\right) = 101.35$, $p < .001$). Specifically, the odds of a participant selecting something other than "strongly disagree" in the first case increased by 4.26 (95% CI [2.95, 6.42]) times for every one-unit higher they scored on the IDRIA ($b = 1.45$, 95% CI [1.08, 1.86], $SE = 0.20$, Wald = 7.30, $p < .001$). The odds of a participant selecting something other than the middle response option in the second case increased by 4.33 (95% CI [3.08, 6.33]) times for every one-unit higher they scored on the IDRIA ($b = 1.47$, 95% CI [1.12, 1.85], $SE = 0.18$, Wald = 7.98, $p < .001$).

Taken together, these findings indicate that, as in the prior studies, the IDRIS and IDRIA are valid predictors of C/IE responding. Nevertheless, there were some notable departures from our hypotheses. First, inconsistent with Hypothesis 3, the IDRIS ($r = .10$, $p = .022$) and IDRIA ($r = .06$, $p = .148$) were not associated with providing long strings of identical responses. This appeared to be due to a markedly low incidence of straightlining in the dataset. A closer inspection of the data revealed that there were, however, a large number of participants who seemed to be cycling through a small set of responses (e.g., "Strongly agree", "Moderately agree", "Slightly agree"), perhaps as a way to avoid being flagged for straightlining. To test the idea that some participants may have adopted a sort of "bandlining" to avoid being detected as straightliners, we recoded any level of disagreement as – 1, "Neither agree nor disagree" as

[18] As in the prior studies, the conclusions remained the same if the data was reanalyzed using (1) Spearman rank-order correlations and (2) Pearson correlations after implementing exponential- and log-based transformations (see the Supplementary Material).

0, and any level of agreement as 1 and reran the long string analyses. The IDRIS ($r = .38$, $p < .001$) and IDRIA ($r = .44$, $p < .001$) were both highly associated with the updated long-string index, suggesting that participants may have, indeed, shifted to a less conspicuous form of perseverative responding.

As a second departure from Hypothesis 3, the IDRIA was not significantly associated with the semantic antonyms index ($r = .12$, $p = .009$). The IDRIS *was* associated with the semantic antonyms index, but the correlation was smaller than that seen in the prior studies ($r = .19$, $p < .001$). These results may be due to the fact that we only used three pairs of items from the Semantic Antonyms Set in the present study instead of the seven pairs of items used in the previous studies. The scale may have simply been a less reliable indicator of inconsistent responding in the present study.

Finally, departing from Hypothesis 6, we found that participants scoring high on the IDRIS $\left(\chi^2(1, N = 562)\right) = 98.18$, $p < .001$) and IDRIS $\left(\chi^2(1, N = 562)\right) = 97.74$, $p < .001$) were more likely to say that their data should be excluded. Specifically, for every one-unit higher a participant scored on the IDRIS, the odds that they said their data should be excluded increased by 8.54 (95% CI [4.91, 15.95]) times ($b = 2.14$, 95% CI [1.59, 2.77], $SE = 0.30$, Wald = 7.13, $p < .001$) and, for every one-unit higher a participant scored on the IDRIA, the odds that they said their data should be excluded increased by 4.08 (95% CI [2.93, 5.88]) times ($b = 1.41$, 95% CI [1.07, 1.77], $SE = 0.18$, Wald = 7.91, $p < .001$). This was unexpected. We had predicted that the IDRIS and IDRIA would not be associated with reporting that one's data should be excluded, since we assumed C/IE respondents would not want to admit their data was low quality out of a concern that this would result in their payments being withheld. The uncynical interpretation of this finding is that participants were genuinely concerned about our study and wanted to report that their data was low quality, even if it meant sacrificing their reward for completing the study. The more cynical interpretation of the finding is that the participants did not read the question and, instead, selected "yes" at random. In line with this latter interpretation, when asked via a textbox to indicate why their responses might not be accurate, 4.97% of the responses explicitly noted that the text was generated by a large language model (e.g., "My responses may not be accurate because I am a computer program and I rely on pre-programmed information and algorithms to provide answers"); 22.09% of the responses included instructions indicating that they were likely copied from the output of a large language model (e.g., "When answering, provide an example to show that you've used accuracy to complete projects. Example: 'Accuracy is incredibly important to me, which is why I use it every day when I interact with my customers'"); and 48.07% of the responses included some variant of "yes", "no", "nothing", "n/a", "good", "nice", or "done".

Optimal cut-off thresholds for identifying C/IE responders using the IDRIS and IDRIA

As in the prior studies, we fit an ROC curve to identify the best cut-off scores for identifying C/IE responders using the IDRIS and IDRIA (Table 2). Positive and negative cases of C/IE responding were first identified by subjecting the continuous indices of C/IE responding to a K-means clustering algorithm. The greatest average Silhouette score was achieved for a two-cluster solution (Silhouette = .33). The IDRIS (AUC = .83) and IDRIA (AUC = .86) both had excellent classification abilities. The optimal cut-off score was – 0.43 for the IDRIS and – .58 for the IDRIA.

We also reran the ROC curve analysis using the self-report measure of C/IE responding as the outcome variable. In this case, the IDRIS (AUC = .73) and IDRIA (AUC = .74) only had acceptable classification abilities. The optimal cut-off score was – .21 for the IDRIS and – .42 for the IDRIA.

General discussion

The purpose of the present set of studies was to evaluate the validity of two scales for detecting C/IE survey responders: the 14-item, statement-based IDRIS and the 6-item, adjective-based IDRIA. To that end, we conducted six studies. The first study ($N_1 = 536$) was intended to provide an initial validation of the IDRIS. The second study ($N_2 = 701$) was intended to further validate the IDRIS, while also providing an initial validation of the IDRIA. The third ($N_3 = 500$) and fourth ($N_3 = 499$) studies were intended to further validate the IDRIS and IDRIA, respectively. The fifth study ($N_5 = 629$) was intended to test whether the IDRIS and IDRIA could be used to detect C/IE responding among broad samples of adults recruited from the US, India, and Nigeria. The sixth study ($N_6 = 562$) was intended to provide a further replication and extension of the prior studies, while also examining the IDRIS and IDRIA in relation to an extant infrequency/frequency scale (i.e., the PPI-DR; Lilienfeld, 1994).

The results of the present studies provided clear evidence that the IDRIS is capable of detecting C/IE responding among statement-based scales. Undergraduate students scoring high on the IDRIS were more likely to speed through surveys, provide long strings of identical responses, exhibit low response variabilities, depart from the average pattern of responses, provide different responses to psychometrically synonymous items, and provide similar responses to psychometrically antonymous items (Study 1, Study 2,

and Study 3). They were also more likely to provide fake e-mail addresses when asked to provide contact information for informants (Study 1); report that their data should be excluded from further analysis for data quality reasons (Study 2); and provide similar responses to semantically antonymous items (Study 3). Additionally, participants sampled from the US, India, and Nigeria who scored high on the IDRIS were more likely to exhibit low response variabilities, depart from the average pattern of responses, provide different responses to psychometrically synonymous items, and provide similar responses to psychometrically antonymous items (Study 5), even when the data was prescreened to remove participants who sped through or straightlined the surveys. Furthermore, participants sampled from MTurk who scored high on the IDRIS were more likely to speed through surveys, exhibit low response variabilities, exhibit low horizontal cursor variabilities, depart from the average pattern of responses, provide different responses to psychometrically synonymous items, provide similar responses to psychometrically antonymous items, and provide similar responses to semantically antonymous items (Study 6). They were also more likely to score high on an existing infrequency/frequency measure (i.e., the PPI-DR; Lilienfeld, 1994), fail instructed response items, and report that their data should be excluded from further analysis for data quality reasons.

The results of the present studies also provided clear evidence that the IDRIA is capable of detecting C/IE responders among adjective-based and adjective-derived scales. Undergraduate students scoring high on the IDRIA were more likely to speed through surveys, provide long strings of identical responses, exhibit low response variabilities, depart from the average pattern of responses, provide different responses to psychometrically synonymous items, and provide similar responses to psychometrically antonymous items (Study 2 and Study 4). They were also more likely to report that their data should be excluded from further analysis for data quality reasons (Study 2); provide similar responses to semantically antonymous items (Study 4); and fail instructed response items (Study 4). Participants sampled from the US, India, and Nigeria who scored high on the IDRIA were also more likely to exhibit low response variabilities, depart from the average pattern of responses, provide different responses to psychometrically synonymous items, and provide similar responses to psychometrically antonymous items (Study 5). Again, this was even when the data was prescreened to remove participants who sped through or straightlined the surveys. Similarly, participants sampled from MTurk who scored high on the IDRIA were more likely to speed through surveys, exhibit low response variabilities, exhibit low horizontal cursor variabilities, depart from the average pattern of responses, provide different responses to psychometrically synonymous items, and provide similar responses to psychometrically antonymous items (Study 6). They were also more likely to score high on an existing infrequency/frequency measure (i.e., the PPI-DR; Lilienfeld, 1994), fail instructed response items, and report that their data should be excluded from further analysis for data quality reasons.

Taken together, the results of the present studies indicate that the IDRIS and IDRIA are capable of detecting C/IE responding. Although more validation work can (and should) be done, we feel comfortable recommending researchers use the IDRIS and IDRIA to detect C/IE responders among statement-based, adjective-based, and adjective-derived scales.

There does, however, remain the issue of what cut-off scores researchers should use to identify C/IE responders when using the IDRIS and IDRIA. Ultimately, the best cut-off scores for a researcher to use will depend on the specific characteristics of their study. Therefore, we generally recommend researchers use a histogram-based approach (see Kay & Saucier, 2023) to identify cut-off scores. Using this approach, researchers produce a histogram of their infrequency/frequency scale scores. They then try to identify the single value that best separates the non-C/IE responders, who should be clustered on the left-hand side of the histogram, from the C/IE responders, who should be clustered on the right-hand side of the histogram.

We do, however, appreciate that researchers may prefer more concrete cut-off thresholds for the IDRIS and IDRIA. To that end, we fit ROC curves to identify optimal cut-off thresholds for identifying C/IE respondents in five of the six studies reported here. The results indicated that, when using a five-point response scale, a cut-off threshold somewhere between – .96 and – .14 would provide the best trade-off between sensitivity and specificity for the IDRIS and a cut-off threshold somewhere between – .75 and – .28 would provide the best trade-off between sensitivity and specificity for the IDRIA. Based on these results, we recommend a cut-off threshold of zero for both scales.

A cut-off threshold of zero differs slightly from the cut-off thresholds recommended by the ROC curves, but we have three reasons for providing this recommendation. The first is that a cut-off threshold of zero makes it slightly harder to incorrectly reject non-C/IE responders. By definition, the optimal cut-off value from an ROC curve balances sensitivity and specificity. However, in this case, it is not clear that we want to be equally likely to exclude non-C/IE responders as we are to include C/IE responders. In fact, the prevailing wisdom among C/IE researchers (e.g., Curran, 2016; Smith & Burger, 1997; Tellegen, 1988) is to err on the side of retaining C/IE responders if it means reducing the number of non-C/IE responders who are incorrectly flagged for exclusion. A cut-off value of zero is a step in that direction, sacrificing some sensitivity in order to obtain greater specificity.

The second reason for our recommendation is that it will result in the exclusion of any participant who, at a minimum, fails approximately half of the infrequency/frequency items from the IDRIS and IDRIA. For example, a participant who selects "strongly agree" to all of the infrequency and frequency items from the IDRIS would be flagged as a C/IE responder, as would a participant who selects "strongly disagree" to all of the infrequency and frequency items from the IDRIS. The reason that this is a desirable property for a cut-off threshold is that it can be fairly easily communicated to various research stakeholders. For example, it makes it easier to explain the cut-off threshold to participants during informed consent and to reviewers during peer review. It also makes it easier to explain the cut-off threshold to administrators at on-demand data collection platforms, which may be necessary to implement the IDRIS and IDRIA as a part of one's exclusionary criteria (see Prolific, 2024).

The final reason we recommend a cut-off threshold of zero is that it avoids researchers having to transform the threshold for every response scale they want to use. A cut-off threshold of – 0.50 on a five-point response scale would, for instance, need to be transformed to be used with a four-point response scale, as it would to be used with a six-point or seven-point response scale. A cut-off threshold of zero remains the same regardless of the response scale used.

Limitations and future directions

The present set of studies had several limitations that are worth noting. First, by necessity, the indices of C/IE responding used here were only a subset of all of the possible indices of C/IE responding that could have been used to validate the IDRIS and IDRIA. We did take care to use a diverse set of C/IE indices, including so-called "direct" (e.g., self-report methods), "archival" (e.g., response durations), and "statistical" (e.g., psychometric synonyms) indices (see Desimone et al., 2015), but other indices exist that could have also been informative (e.g., interactions with approximate areas of interest in surveys; Pokropek et al., 2024). We encourage researchers to use these additional indices to further evaluate the validity of the IDRIS and IDRIA in future projects.

Second, the studies only evaluated the ability for the IDRIS and IDRIA to detect C/IE responding in samples drawn from three countries: the US, India, and Nigeria. Collecting data from these three countries is a good first step in demonstrating the IDRIS and IDRIA's ability to detect C/IE responding across countries, but it in no way ensures the IDRIS and IDRIA would be able to detect C/IE responding in all countries. Further cross-national validation is an important next step for these scales.

Third, we did not examine whether including items from the IDRIS and IDRIA affected participants' subjective experiences completing the surveys. The reason we note this as a potential limitation is that there have been some claims, albeit anecdotal, that infrequency/frequency items can irritate participants. The little empirical work that has been done on the subject has, however, indicated that including infrequency/frequency items in a survey has little impact on a participant's enjoyment of the survey, the participant's belief that the survey is easy to understand, and the participant's willingness to complete future surveys offered by the same researchers (Huang et al., 2015b). In fact, the results suggested that including infrequency/frequency items in a survey actually leads participants to think the results of the study will be of higher quality. We encourage researchers to examine how including the IDRIS and IDRIA items in a survey affects participants' subjective experiences completing the survey, but, given the prior research on the topic, we expect any negative consequences to be minimal.

Fourth, we did not attempt to decrease the length of the IDRIS in the present study, focusing instead on providing a comprehensive investigation into the validity of the full scale. Increased survey length (and a corresponding increase in administration time and cost) is a challenge of using any infrequency/frequency scale, but it is especially an issue for the IDRIS. At 14 items, the IDRIS is longer than many (Beach, 1989; Benning et al., 2018; Huang et al., 2015b; Lilienfeld & Widows, 2005; Lynam et al., 2011; Maniaci & Rogge, 2014; Meade & Craig, 2012) but not all (Ben-Porath & Tellegen, 2020; Dunn et al., 2018; Smith & Burger, 1997) extant infrequency/frequency scales. In some ways, its length is an asset. A longer infrequency/frequency scale should be less likely to incorrectly flag non-C/IE responders (see Footnote 2). Still, if minimizing administration time and cost is a researcher's top priority, such as when prescreening participants, the length of the IDRIS is not ideal. We, therefore, encourage researchers to test the efficacy of shorter versions of the IDRIS in future work. To aid in this effort, we have included the correlations of the individual IDRIS (and IDRIA) items with the indices of C/IE responding in the Supplementary Material.

Finally, we want to emphasize that the IDRIS and IDRIA should not be used to the exclusion of every other method for detecting C/IE responding. All techniques and methods for detecting C/IE responding, including the infrequency/frequency-item method, have their strengths and weaknesses. To assess C/IE responding accurately, it is best to use multiple C/IE-detection methods in concert.

Conclusion

The results of the six studies reported here indicate that the IDRIS and IDRIA are capable of identifying C/IE responders among statement-based, adjective-based, and adjective-derived scales. Although we certainly encourage future work to further validate these two scales, we believe the present results provide good evidence that both scales can be useful tools for screening one's data.

Appendix 1

The Invalid Responding Inventory for Statements (IDRIS)

Directions: Indicate your level of agreement with each of the following statements. Respond to each statement using the following scale.

2 = strongly agree
1 = agree
0 = neither agree nor disagree
−1 = disagree
−2 = strongly disagree

- I have forgotten my last name on multiple occasions.
- If I heard a loud noise behind me, I would turn around to see what it was. (R)
- I sometimes go several weeks without brushing my teeth.
- It should be illegal to intentionally kill an innocent person. (R)
- When someone tells me a funny joke, I often feel angry.
- I try to listen when someone I care about is telling me something. (R)
- I am older than my parents.
- I often say some form of goodbye right before I end a phone call. (R)
- I frequently forget whether my eyes are open or closed.
- When I watch a funny movie, I sometimes smile or laugh. (R)
- I think it should be against the law to listen to music.
- I try to shower or bathe at least once a month. (R)
- When a friend greets me, I generally try to say nothing back.
- I can remember the names of most of my close family members. (R)

Scoring directions. To create an index of C/IE responding, the participants' responses to all of the statements should be averaged together. Statements followed by (R) should be reverse-scored (e.g., -2 becomes 2) prior to averaging.

Appendix 2

The Invalid Responding Inventory for Adjectives (IDRIA)

Directions: Indicate your level of agreement with each of the following statements. Respond to each statement using the following scale.

2 = strongly agree
1 = agree
0 = neither agree nor disagree
−1 = disagree
−2 = strongly disagree

- I am **asleep**, not awake.
- I am **mortal**, able to die. (R)
- I am perfectly **triangular**.
- I am **human**. (R)
- I tend to be **carbonated**.
- I am **literate**, able to read and write. (R)

Scoring directions. Only the bolded adjectives should be administered with adjective-based scales (e.g., the Big Five Mini-Markers; Saucier, 1994). The full IDRIA statements should be administered with adjective-derived scales (e.g., the Big Five Inventory – 2; Soto & John, 2017a). To create an index of C/IE responding, the participants' responses to all of the items should be averaged together. Items followed by (R) should be reverse-scored (e.g., -2 becomes 2) prior to averaging.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-024-02452-x.

Acknowledgements I thank Holly Arrow at the University of Oregon in Eugene, Oregon, and the faculty research grant committee at Union College in Schenectady, New York, for their generosity in funding Study 5 and Study 6, respectively.

Authors' contributions Cameron S. Kay oversaw all aspects of the present manuscript.

Funding Study 5 was funded by departmental funds from Holly Arrow at the University of Oregon in Eugene, Oregon.

Study 6 was funded by a faculty research grant from Union College in Schenectady, New York.

Availability of data and materials The data and materials for all studies are available at https://osf.io/dpckj/?view_only=680406f9cfdb41a3b12006241973def0.

Code availability The R code for Study 6 is available at https://osf.io/dpckj/?view_only=680406f9cfdb41a3b12006241973def0.

Declarations

Conflict of interest/Competing interests The first author of the present manuscript created the tools validated here. There is currently no way for the first author to profit financially from these tools, and he does not intend to ever change this arrangement.

Ethics approval All of the studies reported here were approved or determined to be exempt by the Institutional Review Board at the University of Oregon (Study 1: 10032020.003; Study 2: 8282019.043; Study 3: 09072010.006-MOD00000096; Study 4: STUDY00000083; Study 5: 1122021.014) or the Human Subjects Review Committee at Union College (Study 6: E23027). They were conducted in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent to participate Participants provided informed consent at the beginning of all of the surveys reported here.

Consent for publication Not applicable.

References

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods, 52*(6), 2489–2505.

Ashton, M. C., & Lee, K. (2005). Honesty-Humility, the Big Five, and the Five-Factor Model. *Journal of Personality, 73*(5), 1321–1354.

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2012). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

Beach, D. A. (1989). Identifying the random responder. *Journal of Psychology, 123*(1), 101–103.

Beals, J., Manson, S. M., Mitchell, C. M., Spicer, P., AI-SUPERPFP Team. (2003). Cultural specificity and comparison in psychiatric epidemiology: Walking the tightrope in American Indian research. *Culture, Medicine and Psychiatry, 27*, 259–289.

Benning, S. D., Barchard, K. A., Westfall, R. S., Brouwers, V. P., & Molina, S. M. (2018). Development of the meanness in psychopathy-self report: Factor structure, criterion-related validity, and incremental validity. *PsyArxiv Preprint*. https://doi.org/10.31234/osf.io/8qbgd

Ben-Porath, Y. S., & Tellegen, A. (2020). *Minnesota Multiphasic Personality Inventory-3 (MMPI-3): Manual for administration, scoring, and interpretation*. University of Minnesota Press.

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

Bowling, N. A., Huang, J. L., Brower, C. K., & Bragg, C. B. (2021). The quick and the careless: The construct validity of page time as a measure of insufficient effort responding to surveys. *Organizational Research Methods*, 1–30.

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion, 3*(4), 381–393.

Cheung, F. M., Song, W. Zheng., & Butcher, J. N. (1991). An Infrequency Scale for the Chinese MMPI. *Psychological Assessment, 3*(4), 648–653. https://doi.org/10.1037/1040-3590.3.4.648

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. Academic Press Inc.

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment, 24*(1), 21–35.

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement, 70*(4), 596–612.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*(4), 475–494.

Crowe, M., Carter, N. T., Campbell, W. K., & Miller, J. D. (2016). Validation of the Narcissistic Grandiosity Scale and creation of reduced item variants. *Psychological Assessment, 28*(12), 1550–1560.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4–19.

Curran, P. G., & Hauser, K. A. (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality, 82*, 103849.

Desimone, J. A., Harms, P. D., & Desimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36*, 171–181.

DeSimone, J. A., DeSimone, A. J., Harms, P. D., & Wood, D. (2018). The differential impacts of two forms of insufficient effort responding. *Applied Psychology, 67*(2), 309–338.

Donlon, T. F., & Fischer, F. E. (1968). An index of an individual's agreement with group-determined item difficulties. *Educational and Psychological Measurement, 28*, 105–113.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS One, 18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

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*(1), 105–121.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168.

Gentile, B., Miller, J. D., Hoffman, B. J., Reidy, D. E., Zeichner, A., & Campbell, W. K. (2013). A test of two brief measures of grandiose narcissism: The Narcissistic Personality Inventory-13 and the Narcissistic Personality Inventory -16. *Psychological Assessment, 25*(4), 1120–1136.

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Personality Processes and Individual Differences, 48*(1), 82–98.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*(1), 84–96.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character, 1: Studies in deceit*. The MacMillan Company.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Inventory*. The Psychological Corporation.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–135.

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology, 130*(2), 149–168.

Holtzman, N. S., & Donnellan, M. B. (2017). A simulator of the degree to which random responding leads to biases in the correlations between two individual differences. *Personality and Individual Differences, 114*, 187–192. https://doi.org/10.1016/j.paid.2017.04.013

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*(5), 581–595.

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114.

Huang, J. L., 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*(2), 299–311.

Huang, J. L., Liu, M., & Bowling, N. A. (2015b). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology, 100*(3), 828–845.

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

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). Guilford Press.

Juster, F. T., & Suzman, R. (1995). An overview of the health and retirement study. *Journal of Human Resources, 30*, S7–S56.

Kay, C. S. (2019). *panoply: A Panoply of Miscellaneous Functions*. R Package version 0.0.0.9000.

Kay, C. S. (2021). The targets of all treachery: Delusional ideation, paranoia, and the need for uniqueness as mediators between two forms of narcissism and conspiracy beliefs. *Journal of Research in Personality*, *93*, 104128.

Kay, C. S. (2023). *The anatomy of antagonism: Exploring the relations of 20 lexical factors of personality with Machiavellianism, grandiose narcissism, and psychopathy [Doctoral dissertation]*. University of Oregon.

Kay, C. S., & Saucier, G. (2023). The Comprehensive Infrequency/Frequency Item Repository (CIFR): An online database of items for detecting careless/insufficient-effort responders in survey data. *Personality and Individual Differences*, *205*, 112073.

Kay, C. S., & Slovic, P. (2023). The Generic Conspiracist Beliefs Scale - 5: A short-form measure of conspiracist ideation. *Journal of Research in Personality, 102*, 104315.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42*(9), 1879–1890.

Lachman, M. E., & Weaver, S. L. (1997). *The Midlife Development Inventory (MIDI) personality scales: Scale construction and scoring*. Technical Report, Brandeis University.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329–358.

Lilienfeld, S. O. (1994). Conceptual problems in the assessment of psychopathy. *Clinical Psychology Review, 14*(1), 17–38. https://doi.org/10.1016/0272-7358(94)90046-9

Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathic Personality Inventory - Revised (PPI-R): Professional Manual*. PAR.

Lynam, D. R., Gaughan, E. T., Miller, J. D., Miller, D. J., Mullins-Sweatt, S. N., & Widiger, T. A. (2011). Assessing the basic traits associated with psychopathy: Development and validation of the Elemental Psychopathy Assessment. *Psychological Assessment, 23*(1), 108–124.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*(1), 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Maples-Keller, J. L., Williamson, R. L., Sleep, C. E., Carter, N., Campbell, W. K., & Miller, J. D. (2019). Using item response theory to develop a 60-item representation of the NEO-PI-R using the International Personality Item Pool: Development of the IPIP-NEO-60. *Journal of Personality Assessment, 101*, 4–15.

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.

Mayer, M. (2021). *missRanger: Fast Imputation of Missing Values.* R Package version 2.1.3.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology, 30*(5), 525–564. https://doi.org/10.1037/h0053634

Moss, A. J., & Litman, L. (2018). *After the bot scare: Understanding what's been happening with data collection on MTurk and how to stop it*. CloudResearch. Retrieved April 13, 2024, from https://www.cloudresearch.com/resources/blog/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it/

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867–872.

Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2016). *Self-Report Psychopathy Scale 4th Edition (SRP 4) Manual*. Multi-Health Systems Inc.

Pokropek, A., Żółtak, T., & Muszyński, M. (2023). Mouse chase: Detecting careless and unmotivated responders using cursor movements in web-based surveys. *European Journal of Psychological Assessment, 39*(4), 299–306.

Pokropek, A., Żółtak, T., & Muszyński, M. (2024). Identifying careless responding in web-based surveys: Exploiting sequence data from cursor trajectories and approximate areas of interest. *Zeitschrift Für Psychologie, 232*(2), 95–108.

Prolific (2024). *Prolific's attention and comprehension check policy*. Retrieved April 13, 2024, from https://researcher-help.prolific.com/hc/en-gb/articles/360009223553-Prolific-s-Attention-and-Comprehension-Check-Policy

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*, 590.

Ryff, C. D., Kitayama, S., Karasawa, M., Markus, H., Kawakami, N., & Coe, C. (2018). *Survey of Midlife in Japan (MIDJA-2), May-October 2012 (ICPSR 36427)*. Inter-university Consortium for Political and Social Research. Retrieved April 13, 2024, from https://www.icpsr.umich.edu/web/NACDA/studies/36427

Saucier, G. (1994). Mini-Markers: A brief version of Goldberg's unipolar Big-Five Markers. *Journal of Personality Assessment, 63*(3), 506–516.

Saucier, G., & Iurino, K. (2020). High-dimensionality personality structure in the natural language: Further analyses of classic sets of English-language trait-adjectives. *Journal of Personality and Social Psychology, 199*(5), 1188–1219.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9*(4), 367–373.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*(2), 183–189.

Snyder, C. R., & Fromkin, H. L. (1977). Abnormality as a positive characteristic: The development and validation of a scale measuring need for uniqueness. *Journal of Abnormal Psychology, 86*(5), 518–527.

Soto, C. J., & John, O. P. (2017a). The next Big Five inventory. *Journal of Personality and Social Psychology, 113*(1), 117–143.

Soto, C. J., & John, O. P. (2017b). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69–81.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149.

Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality, 56*(3), 621–663. https://doi.org/10.1111/j.1467-6494.1988.tb00905.x

Thalmayer, A. G., & Saucier, G. (2014). The questionnaire big six in 26 nations: Developing cross-culturally applicable big six, big five and big two inventories. *European Journal of Personality, 28*, 482–496.

Walters, J. (2015). *Qualtrics mouse tracking: A how-to guide for fast implementation*. Retrieved April 13, 2024, from http://math.bu.edu/people/jackwalt/qualtrics-mousetracking/

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*(1), 1–20.

Washburne, J. N. (1935). A test of social adjustment. *Journal of Applied Psychology*, *19*(2). https://doi.org/10.1037/h0053473

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*, 163–183.

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

Yentes, R. D., & Wilhelm, F. (2021). *careless: Procedures for computing indices of careless responding*. R Package version 1.2.2.

Zorowitz, S., Solis, J., Niv, Y., & Bennett, D. (2023). Inattentive responding can induce spurious associations between task behavior and symptom measures. *Nature Human Behaviour, 7*, 1667–1681.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Practices Statement The data and materials for all studies are available at https://osf.io/dpckj/?view_only=680406f9cfdb41a3b12006241973def0.

The R code for Study 6 is available at https://osf.io/dpckj/?view_only=680406f9cfdb41a3b12006241973def0.

Study 6 was the only study that was preregistered: https://osf.io/dpckj/?view_only=680406f9cfdb41a3b12006241973def0.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.