Personality and Individual Differences 84 (2015) 79–83

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

# The *inter-item standard deviation* (ISD): An index that discriminates between conscientious and random responders ☆

Zdravko Marjanovic [a,*], Ronald Holden [b], Ward Struthers [c], Robert Cribbie [c], Esther Greenglass [c]

[a] Thompson Rivers University, Department of Psychology, Kamloops, BC, Canada
[b] Queen's University, Department of Psychology, Kingston, ON, Canada
[c] York University, Department of Psychology, Toronto, ON, Canada

ARTICLE INFO

*Article history:*
Received 14 July 2014
Received in revised form 18 August 2014
Accepted 20 August 2014
Available online 18 September 2014

*Keywords:*
Random responding
Validity scale
Personality
Inventory
Psychometric

ABSTRACT

Although random responding is prevalent and increases Type II errors, most psychologists avoid trying to identify it because the means to do so are extremely limited. We propose the *inter-item standard deviation* (ISD), a statistical index of response variance, is suited for this task. We hypothesized that random responders produce large ISDs because they respond to items all over a measure's response range, whereas conscientious responders produce small ISDs because they respond to items more consistently. We administered a questionnaire containing the NEO-FFI-3 and an embedded validity scale to 134 university students. Another 134 responders were created using a random number generator. For all 268 responders, the ISD was calculated for each of the NEO-FFI-3′s five subscales and an aggregated ISD was calculated by averaging the five ISD indexes. Results showed that (1) random responders produce significantly larger ISDs than conscientious responders, (2) the ISDs were strongly correlated with the embedded validity scale and with one another, and (3) the ISDs correctly identified responders with greater than 80% classification accuracy. The mean ISD yielded greater than 95% classification accuracy. This study shows that responders can be identified by quantifying inter-item response variance.



## 1. Introduction

When we ask individuals to complete psychological inventories,[1] we expect them to follow our instructions and answer all items as honestly and accurately as they can. We expect them to complete the inventories in good faith as *conscientious responders. Random responders* behave oppositely. For a variety of reasons such as carelessness and psychopathology, random responders answer items without regard for what the items mean. For any given item, the random responder is just as likely to give a response at the bottom of a response scale (e.g., Strongly Disagree) as they are to respond at the top (e.g., Strongly Agree) or middle (e.g., Neutral). Their data are completely meaningless and invalid. The problem for test administrators is twofold. First, because random responders consistently produce mean scores around the midpoint of a measure's response range, failure to identify and remove their data reduces the mean score variance in otherwise validate data (Holden, Wheeler, & Marjanovic, 2012). This reduced variance decreases a study's power and increases the likelihood that psychologists make Type II errors (Credé, 2010; Osborne & Blanchard, 2011). This is especially worrying considering that the prevalence of random responding is around 10% across disordered and non-disordered populations (Meade & Craig, 2012).

The second problem is that random responders are notoriously difficult to detect. A recent evaluation of the different available tools for identifying random responders revealed that few tools are effective on their own (Meade & Craig, 2012) To increase one's chances at success, the authors suggested using multiple indices in a type of shotgun or combined-model approach, which hardly inspires confidence. A similarly disappointing conclusion was reached by Huang, Curran, Keeney, Poposki, and DeShon (2012). In their review, the best three indices were effective at correctly classifying conscientious responders as conscientious, producing greater than 90% accuracy, but all three failed to adequately classify random responders as such, producing accuracies well below .50.

Historically, the only effective tools for discriminating between random and conscientious responders have been standardized infrequency scales and inconsistency scales, the kind that are

☆ We thank Lisa M. Bajkov for her assistance with this research.
* Corresponding author.
*E-mail address:* zmarjanovic@tru.ca (Z. Marjanovic).
[1] A psychological measure in which examinees' item responses cannot be evaluated objectively as to whether they are correct or incorrect. Instead, responses reflect the extent to which a responder agrees with some statement or believes it applies to them personally.

http://dx.doi.org/10.1016/j.paid.2014.08.021
0191-8869/

found in multidimensional batteries such as the MMPI and PAI series. Unfortunately, in spite of their effectiveness, the shortcomings associated with these scales are substantial enough that psychologists do not employ them. Their main drawback is that they require a great deal of resources to develop, validate, and norm. For most psychologists' purposes, this makes them more trouble than they are worth. Heretofore, it has been simpler to just accept the variance-killing effect of random responding than to try to do anything about it.

### 1.1. Instructional item content

Marjanovic, Struthers, Cribbie, and Greenglass (2014) attempted to remedy the situation by introducing a novel tool for identifying conscientious responders in inventory data: the five-item *Conscientious Responders Scale* (CRS). Its advantages over traditional random responding scales stem from its use of instructional item content. Because CRS items instruct responders exactly how to answer each item, there is only one expected way to respond – compliantly. Item responses can therefore be scored objectively as either compliant (scored as a 1) or incompliant (scored as a 0), which is impossible to do with traditional random responding scales. Given that instructional items can be scored objectively, one needs only to use probability theory to generate expected rates of item responses. Normative testing is not required. For example, on a seven-point Likert scale, we know that a random responder is less than 5% likely to answer three items compliantly out of five by chance alone. Using the ubiquitous .05 *p*-value as our guide, we selected 2/3 as our optimal cutoff. That is, responders who answer 0, 1, or 2 items compliantly are labelled "random responder" because their scores are indistinguishable from what can be achieved by random number generator, and scorers of 3, 4, or 5 are labelled "conscientious responder" because of the improbability that random responders could score so high.

The results of Marjanovic et al. (2014) demonstrated that the CRS is effective at discriminating between responder groups. In two experiments in which the CRS was randomly distributed throughout the length of a questionnaire, the CRS correctly discriminated between conscientious and random responders with about 93% accuracy. In addition, when we compared the classification accuracy of the theoretically-generated 2/3 cutoff against an empirical cutoff generated from the data itself, results revealed they were identical. This was not the case when we examined the classification accuracy of a traditional infrequency scale. Its optimal cutoff was different across the two studies. This is an advantage for the CRS or for any validity measure containing instructional items. One limitation of the CRS, however, which similarly affects infrequency scales, is that it has to be embedded in a questionnaire before the questionnaire is administered. It cannot be employed afterward as a type of *post-hoc* afterthought. So in spite of its effectiveness, this raises the question; what is a researcher to do if they did not consider random responding before they collected their data, but suddenly wants to screen for it now? Is it even possible with existing methods?

In short: no. Several methods for identifying random responders have been developed and tested that do not depend on embedded scales. For example, the long string approach flags *hyper-consistent responders* (i.e., individuals who endorse too many items similarly in a row), the completion time approach flags responders who complete a questionnaire too quickly to be valid, and the consistency approach flags responders who respond inconsistently to pairs of items that are overwhelmingly responded to consistently by others (Huang et al., 2012). Thus, long strings, brief completion times, and low response consistency are thought to indicate random responding. The trouble with these indexes and others like them is that they work so dismally that no one bothers using them (Meade & Craig, 2012). In practice, there are no effective *post-hoc* means for psychologists to identify responders in inventory data.

## 2. The present study: the response variance hypothesis

The purpose of this investigation was to test a statistical approach for discriminating between responders which is not dependent on embedded scales. We propose that the *inter-item standard deviation* (ISD)[2] is index suited to this task. Similar to the familiar *standard deviation* (SD), which is an interpersonal measure of variability calculated at the group level, the ISD is an intrapersonal measure of response variance calculated at the individual level. On any given measure, the ISD reflects how closely a responder's item responses cluster around his/her composite mean score. Because we expect that conscientious responders respond similarly to items that tap into the same psychological construct and random responders respond widely across the breadth of the response scale, conscientious responders should produce small ISD scores whereas random responders produce large ISD scores. We hypothesized that the magnitude of the difference in response variance between these responders is large enough to make it useful for identifying them.[3] The following six hypotheses were tested.

1. Conscientious responders produce larger CRS scores than random responders.
2. Conscientious responders produce smaller ISD scores than random responders.
3. The ISD would demonstrate its validity by negatively correlating with the CRS.
4. Based on its theoretically derived cutoff, the CRS would classify participants in the conscientious responder group as conscientious and participants in the random responder group as random with ⩾80% classification accuracy (c.f., Clark, Gironda, & Young, 2003).
5. Based on empirically derived cutoff scores, the ISD indexes would also be able to discriminate between responders group at ⩾80% accuracy.
6. Consistent with the principle of aggregation, a mean ISD score, which is aggregated from several unidimensional ISD scores, would produce a more precise, reliable index of response variance than the unidimensional ISDs that make it up.

## 2. Method

### 2.1. Participants

The *Conscientious Responder* (CR) group consisted of 134 undergraduate students who were recruited from two introductory psychology classes in exchange for course credit. They consisted of 86 women, 46 men, and had a mean age of 19.80 ($SD = 3.54$). Two

[2] The formula for ISD is identical to the unbiased estimate of standard deviation except that a single respondent's mean score is used in the place of the group's mean score. The ISD is:

$$\mathrm{ISD}_j = \sqrt{\frac{\sum_{j=1}^{k}\left(X_j - \overline{X}_i\right)^2}{(k-1)}}$$

where $X_j$ = a respondent's item score, $X_i$ = a respondent's mean score across all scale items, and $k$ = total number of scale items. Before calculating the ISD, be sure that all items (1) measure the same psychological construct and (2) are positively correlated with one another (i.e., all negatively-worded items have been reverse-scored).

[3] With few exceptions (e.g., Farrell, Danish, & Howard, 1999), the ISD or similar response variance indexes have not yet been used to identify random responding. They have, however, been used in related psychometric fields that assess item-response accuracy (Charter, 2000), personality stability (Asendorpf, 1992), and test reliability (Sturman, Cribbie, & Flett, 2009).

participants did not report their sex. The *Random Responder* (RR) group consisted of 134 participants that were created using a random number generator (Haahr, 2013). The total sample contained 268 responders.

### 2.2. Measures

*CRS* (Marjanovic et al., 2014). The CRS is a 5-item validity scale that discriminates between conscientious and random responders in self-report inventory data. Each of its items instructs responders exactly how to answer that item (e.g., Item 2. "Choose the first option – 'strongly disagree' – in answering this question"). A conscientious responder, because he/she is paying attention to item content, is expected to answer all CRS items compliantly, which are scored as 1s. A random responder, because he/she is not paying attention to item content, is expected to answer most CRS items incompliantly, which are scored as 0s. The random responder's probability of answering items compliantly by chance alone = 1/# of response options. Item scores are then summed to create a total CRS score that ranges between 0 and 5.

In this study the CRS was completed using a 5-point Likert scale (1 = *Strongly Disagree* to 5 = *Strongly Agree*), thus the probability that a random responder could correctly answer an item was 20%. With a total of 5 items, probability theory tells us that fewer than 6% of random responders could achieve a score of 3 or more.[4] Given this, we labelled any responder who produced a low CRS score of 0, 1, or 2 as a "random responder" and a high score of 3, 4, or 5 as a "conscientious responder."

*NEO-Five-Factor Inventory-3* (NEO-FFI-3; McCrae & Costa, 2010). This NEO-FFI-3 is a short form of the longer NEO-PI-3. Its 60 items are divided equally among 5 subscales, each tapping into one dimension of the five-factor personality model: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Its items are answered on a 5-point Likert scale ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree*. It has excellent psychometric properties and takes about seven minutes to complete. A sample item is, "I am a very active person."

### 2.3. Procedure

To collect the CR data, our questionnaire was group administered in two introductory psychology classes. The questionnaire contained the NEO-FFI-3, two demographic items on sex and age, and the CRS which was randomly imbedded throughout the length of the questionnaire. Students were asked to complete the questionnaire using standard testing instructions with an additional blurb that prepared them for the instructional nature of the CRS ("Some of these items will ask you to answer them in a very particular way, so be sure to read all of the items carefully and answer them as honestly and accurately as you can"). After the CR data were collected, the RR group was created using Haahr's, (2013) random number generator, which is consistent with other research in the field of random responding (Clark et al., 2003; Holden et al., 2012). For every item in the questionnaire, a random integer between 1 and 5 was generated and added to the data file, doubling the size of the original sample.

## 3. Results

We began by examining CRS scores for differences across the responder groups. Similar to Marjanovic et al. (2014), the CR group ($M$ = 4.58, $SD$ = 1.16) produced significantly larger means than the RR group ($M$ = 1.09, $SD$ = 0.94), $t(266) = 27.12$, $p < .001$, $d = 3.31$ (Table 1). Hypothesis 1 was therefore supported. The large effect size suggests that typical CRS differences between conscientious and random responders are discernable by eye. When we examined the ISD differences across groups, we also found support for hypothesis 2. Random responders produced larger ISDs than conscientious responders on all five indexes. The largest effect was found with the mean ISD.

Secondly, we examined correlations which could demonstrate the ISD's validity (Table 2). Correlations between the CRS and unidimensional ISDs ranged between −.62 and −.71, with a mean of −.67. As expected, the CRS correlated most strongly with the mean ISD at −.81 due to its increased precision over the unidimensional indexes. The unidimensional ISDs also correlated with one another strongly and in a positive direction ($M$r = .60), showing that they were all likely tapping into the same construct. These findings support hypothesis 3, that ISD scores would consistently reflect levels of responder conscientiousness.

Lastly, we examined how accurately the ISD would classify responders across the two groups. For the sake of comparison, the CRS data was analyzed first. Classification were made based on the CRS' theoretically derived 2/3 cutoff score. Results showed that the classification accuracy of the CRS was highly similar to the results of Marjanovic et al. (2014) and met our expectations for these data, supporting hypothesis 4. Of the 134 participants in the CR condition, 123 produced large enough CRS scores ($\geqslant 3$) to be correctly labelled as "conscientious responder." This represents a sensitivity score of 91.79%. Of the 134 participants in the RR condition, 15 produced large enough CRS scores to be incorrectly labelled as "conscientious responder." This represents a specificity score of 88.81%. Combined, the CRS produced an average classification accuracy of 90.30%.

A logistic regression analysis was then conducted to determine whether a theoretically derived cutoff score was equivalent in its accuracy to an empirically derived one (Table 3). Using the CRS as the predictor variable and responder group as the criterion variable (0 = CR vs. 1 = RR), our analysis produced classifications based on a 50% cut rate. This means that when the estimated probability of a given responder belonging to the RR group met or exceeded 50%, the responder was labelled "random responder." Below 50%, the responder was labelled "conscientious responder."

Empirically derived cutoff scores are presented in Table 3. Results showed that the optimal empirically-derived and theoretically-derived CRS cutoff scores were identical: 2/3. This means that empirically, as well as theoretically, responders with scores $\geqslant 3$ had a greater than 50% probability of being conscientious responders. Thus, both approaches to classifying responders lead to equal rates of sensitivity, specificity, and average classification accuracy.

Next, we conducted a series of similar logistic regressions using the ISD as the predictor variable (Table 3). Note, however, that because the CRS and ISD indexes are negatively related, a participant who had an ISD score $\geqslant$ the 50% cut value was labelled "random responder," not "conscientious responder." The first five logistic regressions tested the classification accuracies of the unidimensional ISDs, whereas the sixth regression tested the mean ISD.

The finding showed that all five ISDs exceeded our 80% classification standard, ranging from 80.37% to 88.89% ($M$ = 85.48%). The sixth regression revealed that aggregating ISDs improved their classification accuracy. The mean ISD's accuracy beat all of the other indexes, including the CRS, approaching 96% classification accuracy. These findings therefore supported hypothesis 5 and 6. On their own, a unidimensional ISD index gives a tester a better than reliable chance to guess whether data are conscientiously or randomly generated, but a tester can increase their chances of classifying correctly by simply aggregating multiple ISDs into a single score.

[4] In Marjanovic et al. (2014) the CRS was administered using a 7-point Likert scale. Because in this study we used a 5-point scale, some of the descriptive wording in the items had to be adjusted to reflect the smaller response scale.

**Table 1**
Descriptive statistics and between-group analyses.

| Measure | Responder group | | | | | Analysis | | | |
|---|---|---|---|---|---|---|---|---|---|
| | CR | | | RR | | Independent samples *t*-tests | | | |
| | ($n$ = 134) | | | ($n$ = 134) | | | | | |
| | *M* | *SD* | α | *M* | *SD* | *t* | df | *p* | *d* |
| CRS | 4.58 | 1.16 | .89 | 1.09 | 0.94 | 27.12 | 266 | <.001 | 3.31 |
| N | 2.96 | 0.67 | .85 | 2.94 | 0.36 | 0.29 | 209.92 | =.770 | 0.04 |
| E | 3.61 | 0.58 | .83 | 3.07 | 0.46 | 8.51 | 253.22 | <.001 | 1.04 |
| O | 3.64 | 0.53 | .78 | 2.87 | 0.37 | 13.81 | 239.98 | <.001 | 1.69 |
| A | 3.61 | .53 | .76 | 3.07 | 0.36 | 9.74 | 236.03 | <.001 | 1.19 |
| C | 3.49 | 0.55 | .83 | 2.97 | 0.43 | 8.64 | 251.07 | <.001 | 1.06 |
| N_ISD | 0.95 | 0.25 | – | 1.41 | 0.17 | −17.43 | 230.82 | <.001 | −2.13 |
| E_ISD | 0.86 | 0.27 | – | 1.40 | 0.18 | −19.22 | 229.20 | <.001 | −2.35 |
| O_ISD | 0.89 | 0.29 | – | 1.43 | 0.18 | −18.18 | 225.18 | <.001 | −2.22 |
| A_ISD | 0.97 | 0.28 | – | 1.39 | 0.23 | −13.93 | 257.42 | <.001 | −1.70 |
| C_ISD | 0.91 | 0.28 | – | 1.35 | 0.19 | −15.03 | 234.60 | <.001 | −1.84 |
| *M*_ISD | 0.92 | 0.18 | – | 1.40 | 0.09 | −27.58 | 194.87 | <.001 | −3.37 |

**Table 2**
Correlations between the CRS and all ISD indexes ($N$ = 268).

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. CRS | | | | | | | |
| 2. N_ISD | −.69 | | | | | | |
| 3. E_ISD | −.71 | .68 | | | | | |
| 4. O_ISD | −.71 | .65 | .60 | | | | |
| 5. A_ISD | −.62 | .64 | .54 | .60 | | | |
| 6. C_ISD | −.62 | .59 | .66 | .49 | .59 | | |
| 7. *M*_ISD | −.81 | .86 | .85 | .81 | .81 | .80 | |

*Note*. All correlations were statistically significant at $p < .001$.

## 4. Discussion

Random responding in inventory data is problematic because (1) it is prevalent, (2) it increases Type II error, and (3) it is difficult to detect. The purpose of this study was to test a statistical means for identifying responders in inventory data called the ISD. We administered a widely used personality inventory to university students and calculated ISD indexes for each of its five subscales. Additionally, we speculated that an aggregated ISD would be a better indicator of responding than any unidimensional index because of the principle of aggregation: the aggregate being more precise, reliable than the indexes that make it up. Thus, a sixth ISD index was calculated by averaging across the five subscale ISDs. Lastly, we embedded the CRS in the questionnaire in order to provide a standard against which to evaluate the ISD's performance.

The findings of this investigation were fully supportive of our response variance hypothesis. Firstly, the ISD scores of conscientious responders were much smaller than those of random responders. Conscientiously-generated responses were highly similar to one another, whereas randomly-generated response answers were all over the response scale. This was true for all five subscales, producing similar results across the five indexes. Secondly, ISD indexes were strongly negatively related to the CRS and strongly related with one another in a positive direction. This supports the validity of the ISD as a reliable indicator of response conscientiousness. Finally, all five unidimensional ISDs exceeded a classification accuracy rate of 80%. With its increased precision, however, the mean ISD index performed best, correctly classifying responders about 96% of the time, even better than the embedded random responding scale. In sum, this study showed that the ISD is an effective indicator of responder conscientiousness and merits additional study as an *ad hoc* means to identify responders in data.

For researchers wishing to experiment with the ISD in their own data, we recommend following the same four step process we used. (1) Generate random data for all of the items in your questionnaire, matching the size of your original sample. Label responders as 1s and 0s as per group membership. (2) Calculate the ISD for each responder in the total sample ensuring that all items in the calculation are highly, positively correlated. (3) Perform a logistic regression analysis using the ISD as the predictor variable and responder group as the criterion variable. This will generate probabilities of group membership for each responder in the total sample. (4) If the average classification accuracy of the regression model is acceptable (e.g., ⩾80% or ⩾90%), expunge all individuals labelled "random responder" from the original human sample. Done in this study using our most reliable indicator, the mean ISD, we would have expunged 8 responders from our human sample, constituting a random responding rate of ≈6%.

### 4.1. Limitations and future research

First, this study's conclusions are based on a single sample, using data collected on a single, albeit highly regarded, personality inventory. The tenability of the response variance hypothesis

**Table 3**
Logistic regression results for the CRS and all ISD indexes ($N$ = 268).

| Variables | $\chi^2$ | df | Sensitivity | Specificity | Classification accuracy | Empirical cutoff |
|---|---|---|---|---|---|---|
| CRS | 251.48 | 1 | 91.79 | 88.81 | 90.30 | 3 |
| N_ISD | 194.28 | 1 | 82.96 | 88.15 | 85.56 | 1.21 |
| E_ISD | 156.45 | 1 | 86.67 | 91.11 | 88.89 | 1.16 |
| O_ISD | 204.67 | 1 | 85.16 | 92.59 | 88.89 | 1.21 |
| A_ISD | 132.41 | 1 | 76.30 | 84.44 | 80.37 | 1.22 |
| C_ISD | 159.08 | 1 | 82.96 | 84.44 | 83.70 | 1.16 |
| *M*_ISD | 290.49 | 1 | 94.03 | 97.76 | 95.90 | 1.22 |

*Note*. All regression models were statistically significant at $p < .001$. Sensitivity = % of correct classifications in the CR group; specificity = % of correct classifications in the RR group. Empirically derived cutoff (for the CRS) = an equal or greater score means the probability of being a "conscientious responder" was greater than 50%. Empirically derived cutoff (for the ISD score) = an equal or greater ISD means the probability of being a "random responder" was greater than 50%.

should be further evaluated using a wider variety of inventories, test settings, and methodologies. The ISD should also be evaluated using human-generated random data, in addition to artificially generated random data as was done here. Although it stands to reason that both of these groups' data are similarly invalid and meaningless, replicating these findings with an all human random sample would more compellingly demonstrate the ISD's ecological validity. Second, the ISD's effectiveness will vary to some extent with a measure's (1) internal consistency, (2) number of items, and (3) response scale size. The greater the number of any of these three variables should relate to smaller ISDs in conscientious responders, which in turn should make it easier to discriminate between responders. We recommend in any future participant-based or simulation studies of the ISD's effectiveness, measures with varying levels of these three factors be used.

Third, if researchers are to employ the ISD, CRS, or similar random responding indicators, they should be prepared for the inevitability of finding random responders. This means expecting that your sample will shrink as you identify and expunge random responders. To prevent inflating Type I and II error rates as you eliminate participants, collect a sample that is at least 10% larger than your minimum requirements. Choosing appropriate cutoff scores for discriminating between responders is a more difficult consideration and depends largely on the potential consequences of the research being conducted. For the purposes of most psychological research, we recommend selecting a cut rate that best approximates standard *p*-critical values, such as .05, and, as much as possible commit equal rates of Type I and Type II errors. Achieving high average classification accuracies is the main goal.

Lastly, for the purposes of this preliminary study, we evaluated the effectiveness of the ISD in a worst-case scenario, trying to identify full-on random responders and their uniformly distributed random data. In this type of scenario the ISD performed admirably. It behooves us to note, however, that people do not only respond in a clean and dichotomous, conscientious vs. random manner, and "random" data are not always uniformly distributed. A whole range of response styles and distributions lie in between. For example, previous research has suggested that *intermittent* or *partial random responding* (i.e., randomly responding to some, but not all of a questionnaire's items) may constitute a larger problem for examiners as it is more prevalent than full-on random responding. It has yet to be determined how much random responding is necessary in a data set for the ISD to reliably discriminate between responders.

Regarding distributions, the term "random responder" has come to symbolize a variety of meaningless, invalid data that are not necessarily randomly distributed. In this way, the use of the term is a misnomer and researchers have tacitly voiced their dissatisfaction with it through their use of alternative labelling, such as careless, indiscriminant, and content nonresponsive responding. One source of random data that does not produce a uniform distribution is hyper-consistent responding. Notably, using the approach we used in this paper in which small ISDs indicated conscientious responding and large ISDs indicated random responding, identifying these responders would be impossible if their responses were near a measure's midpoint and the measure contained only positively-worded items (i.e., no reverse-scored items). In this scenario, these responders would produce very small ISDs and be misclassified as conscientious responders. An alternative approach, which could identify them as well as traditional random responders, would be to (1) identify confidence intervals for conscientiously-generated ISDs (e.g., ±2 SDs) and subsequently (2) label responders within in the interval as "conscientious," below it as "hyper-consistent," and above it as "random." A confidence interval approach would also make it possible to identify responders with ISDs too large to be produced by random responding: for example, individuals who purposefully try to contaminate data by responding inconsistently.

## 5. Conclusion

The *inter-item standard deviation* (ISD) is a new statistical means for psychologists to identify random responders. It does not depend on embedded validity measures and can be calculated in any set of inventory data, old or new. If calculated from a unidimensional scale, the ISD is capable of discriminating between random and conscientious responders with high classification accuracy. This accuracy improves when using an aggregated ISD, which produces classification accuracies equivalent to that of an embedded random responding scale. Together, these findings show that the response variance hypothesis has merit and is a viable option for identifying responders.

## References


Asendorpf, J. B. (1992). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality, 6*, 103–117.

Charter, R. A. (2000). Determining random responding to objective tests. *Journal of Psychoeducational Assessment, 18*, 308–315.

Clark, M. E., Gironda, R. J., & Young, R. W. (2003). Detection of back random responding: Effectiveness of MMPI-2 and personality assessment inventory validity indices. *Psychological Assessment, 15*, 223–234.

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

Farrell, A. D., Danish, S. J., & Howard, C. W. (1999). Evaluation of data screening methods in surveys of adolescents' drug use. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 295–298.

Haahr, M. (2013). *Random integer generator. Random.org.* <http://www.random.org/integers/> accessed 09.12.2013.

Holden, R. R., Wheeler, S., & Marjanovic, Z. (2012). When does random responding distort self-report personality assessment? An example with the NEO-PI-R. *Personality and Individual Differences, 52*, 15–20.

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

Osborne, J. W., & Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science results. *Frontiers in Psychology*. http://dx.doi.org/10.3389/fpsyg.2010.00220.

Marjanovic, Z., Struthers, C. W., Cribbie, R. A., & Greenglass, E. R. (2014). The conscientious responders scale: A new tool for discriminating between conscientious and random responders. *SAGE Open, 4*(3), 1–10. http://dx.doi.org/10.1177/2158244014545964.

McCrae, R. R., & Costa, P. T. Jr., (2010). *NEO inventories for the NEO Personality Inventory-3 (NEO-PI-3), NEO-Five-Factor Invetory-3 (NEO-FFI-3), and NEO Personality Inventory-Revised (NEO-PI-R): Professional manual*. Lutz, FL: PAR.

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

Sturman, E. D., Cribbie, R. A., & Flett, G. L. (2009). The average distance between item values: A novel approach for estimating internal consistency. *Journal of Psychoeducational Assessment, 27*, 409–420.