## Method

Participants: Approximately 482 undergraduate students from a midsize southwestern university (approximately 75% female, 25% male; 95% Hispanic), separated into two different phases. Phase 1 (February-July, 2012): N=286, 75% female. Phase 2 (August-December, 2012): N=196, 81% female.

Procedures & Materials: participants completed an anonymous online survey that included the Big Five Inventory (BFI; John & Srivastava, 1999) and 30 pairs of items selected from the International Personality Item Pool (Goldberg et al., 2006), judged to be semantically related. The degree of discrepancy in participants' responses to these pairs was taken as an indicator of RR.

Experimental Manipulation: In some research sessions (Q, or "quick" condition) research assistants subtly encouraged participants to complete the task quickly and in-test messages emphasized the importance of students' time. In the control (A or "accurate") condition participants were instructed to complete the survey accurately.

SIS Item Selection and Scoring: Selection and validation of final item pairs for the SIS was conducted on Phase 1 data only, and was based on maximizing observed correlations between semantic-pair discrepancy scores and time to complete the full assessment, differences in pair discrepancy scores between the Q and A conditions, and correlations between discrepancy pairs. The resulting 22-item (11-pair) SIS scale was assessed using Phase 2 responses.

## Discussion

The Semantic Inconsistency Scale (SIS), a 22-item (11-pair) measure of random responding, was developed in the Phase 1 sample and validated in Phase 2. Evidence for the validity of the SIS for detecting random responding appears good: it has excellent ability to discriminate between true responses and 100% random responses, and fair performance even with protocols having less than 20% random responding. The SIS also performed relatively well on a more subtle test—the ability to discriminate between responses from participants primed and instructed to answer hastily and control participants.

The SIS appears to perform its assigned task well, and at least as well as (if not better than) comparable tests developed in previous research, many of which are only legally available with expensive commercially-distributed personality assessments. Due to the nature of the items (taken from the IPIP), the SIS is easily inserted into a variety of psychological and personality tests; modification of item stems or formats may allow use with an even wider range.

**Limitations and Future Directions:** The SIS is not appropriate for all test varieties, and its length may preclude its use in very short research or clinical protocols.

## Conclusion

The SIS, available without cost from the authors (Creative Commons licensed), is potentially useful for researchers and clinicians using questionnaires without built-in validity scales. Because the SIS was able to detect both computer-generated random responses and invalid responding caused by more realistic conditions, it appears at this point to be a robust and valid measure of random responding.

### References

Available in handouts, by contacting the authors, or at the URL/QR

http://www.darrinlrogers.com/dissemination/

# Development of a Public-Domain Measure of Random Responding

Emanuel Cervantes,
Sarah E. Stegall,
Darrin L. Rogers

## Abstract

The Semantic Inconsistency Scale (SIS), a no-cost tool for measuring random responding in questionnaire research, was developed and validated in two independent samples. It shows strong initial evidence of validity, able to not only detect computer-generated random responses but also invalid responding caused by more realistic conditions.

## Introduction

Invalid responding to questionnaire-based measures can threaten validity and interpretation in many research and clinical situations (Huang, 2012; however, see Costa & McCrae, 1997 for alternative views). Random responding (RR; Archer & Smith, 2008) has been successfully assessed with scales that measure participants' consistency of to pairs of items with similar—or opposite—meanings (e.g., MMPI2 scale VRIN; Butcher et al., 2001; PAI scale INC; Morey, 2007).

Methods previously used to develop and evaluate RR scales include comparing responses from participants instructed to answer questionnaires randomly with subjects given standard instructions (Berry et al., 1991; Cramer, 1995; Galen & Berry, 1996) and comparing real responses with computer-generated random responses (Charter & Lopez, 2003). Real-world tests using more ecologically valid manipulations to influence participants' response styles might be more externally valid than these methods, but they have apparently not been used so far.

Despite the apparent benefits of validated scales to detect RR, we are unaware of any such scales available to professionals except those tied to commercially-marketed assessments. We will describe the development of a public-domain measure of RR for use with questionnaires: the semantic inconsistency scale (SIS).

## Analyses & Results

**Q versus A comparison**

Phase 2 SIS median scores were significantly higher for the Q than the A condition (Wilcoxon test $z$=2.179, $p$<.05; see Figure 1).

**Survey completion time**

In Phase 2, the correlation between SIS scores and time to complete the full research survey was negative and marginally significant (Spearman's rho = -.13, $p$ = .06).

**Attention to survey content**

In Phase 2, we included questions about the content of survey items they had just seen and responded to. There was no association between SIS scores and accuracy of identifying just-seen content (Spearman's rho = .04, $p$ > .05).

**Discrimination with 100% random responding**

The SIS's ability to discriminate between true random and systematic responding was tested by comparing true Phase 2 responses to 100,000 records of randomly-generated responses. Mean SIS scores were significantly lower than the random responses ($t$=31.56, $p$<.001; Figure 2).

**ROC analysis: True responses vs. 100% random profiles**

Phase 2 responses were compared to a randomly-selected subset of the random responses mentioned above, and a receiver-operator characteristic (ROC) analysis performed. Area under the curve (AUC) for this analysis was over .94 (Figure 3), indicating excellent discrimination between random and actual responses.

**Discrimination between true responses and 1%-100% random profiles**

The Phase 2 dataset was split in half randomly, with one half of participants having a randomly-selected proportion (from 1% to 100%, in turn) of their actual responses to the survey replaced by random simulation responses. At each point the SIS was scored and AUC calculated for discriminating the fully-original from the partially-random records. This process was repeated 100 times. Figure 4 shows AUC (i.e., discriminating power of the SIS) for 1% to 100% "injection" of random responding. The SIS performed well with about 30% random responding and excellently with 45-60% randomness.

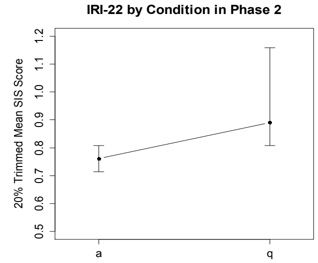


Figure 1. Trimmed (20%) means for SIS scores in condition A ("accurate") versus Q ("quick").

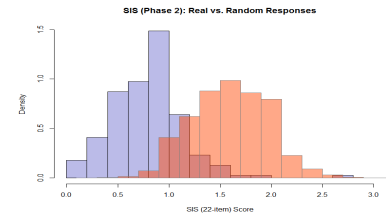


Figure 2. Distribution of true Phase 2 SIS scores (blue) versus randomly-generated profiles (red).

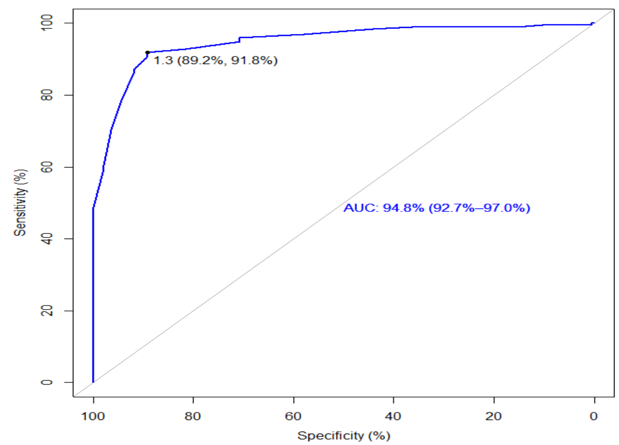


Figure 3. ROC analysis for Phase 2 responses vs. (100%) randomly-generated response records.

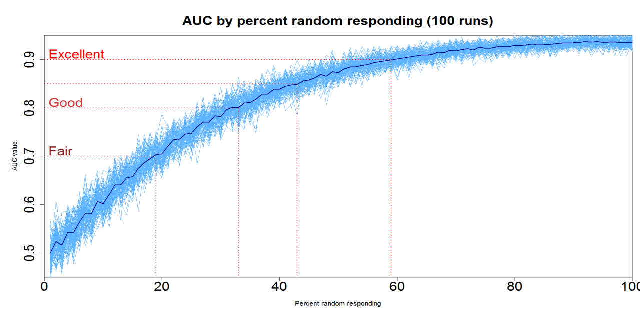


Figure 4. AUCs for 100 runs of SIS discrimination between original