Personality and Individual Differences 129 (2018) 83–87

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Examination of the validity of instructed response items in identifying careless respondents

Chester Chun Seng Kam[a,*], Gabriel Hoi-huen Chan[b]

[a] University of Macau, Macau
[b] Hong Kong Community College, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Keywords:
Careless responding
Instructed response items
Validity

ABSTRACT

This study asks whether instructed response items validly detect careless respondents. We found that instructed response items have high internal consistency, good convergent validity with other types of carelessness detectors (synonyms, antonyms, and odd-even index), and good discriminant validity with acquiescence. Excluding careless respondents based on increasingly strict cutoff values leads to stronger negative correlations between regular and reverse-keyed items. Overall, the results favor the continued use of instructed response items to identify careless respondents.

1. Introduction

Detection of careless responding is often overlooked by personality researchers. It was once assumed that careless responding weakens relationships among constructs, making statistically significant results more conservative, and thus does not greatly affect conclusions based on such findings. Recent research has challenged this conclusion. It is now known that careless responding can inflate or deflate construct correlations based on the type of careless respondent (Kam & Meyer, 2015) and scale means (Huang, Liu, & Bowling, 2015). Therefore, the effect of careless responding can be unpredictable, making its detection important for drawing valid research conclusions.

The purpose of the current study is to investigate the validity of one common type of careless responding, instructed response items. These items direct participants to interact with the survey in a specific way (e.g., "Please choose *strongly disagree* for this item" or "Skip this item and do not choose any response"). Although some research has recommended the use of instructed response items for detecting careless responding, other research has challenged their validity. However, to our knowledge, there has not been systematic investigation of the validity of this method for identifying careless respondents. The investigation is timely because instructed response items have already been widely employed and they will likely continue to be. If this method is problematic, the substantive conclusions made in such investigations would be compromised.

1.1. Previous research on the validity of instructed response items

As mentioned, instructed response items request participants to interact with the survey in a specific manner, such as asking them to select a particular option or skip an item. Researchers may include one or a few such items in a survey. Participants who follow the instructions of the items are deemed careful respondents; those who do not are deemed careless. Researchers, nevertheless, must set a cutoff score; often, they exclude participants who fail to correctly answer one or no items (e.g., Kam & Meyer, 2015; McGonagle, Huang, & Walsh, 2016).

Some researchers have found good validity for these types of instructed response items. Kam and Meyer (2015) and Maniaci and Rogge (2014) identified careless respondents using latent class analysis, and discovered that among various indices, instructed response items were the best for differentiating careful and careless respondents. Specifically, the effect size for differentiating between the two groups was largest for instructed response items, followed by longstring and Mahalanobis distance scores (Kam & Meyer, 2015; Maniaci & Rogge, 2014); total survey minutes were relatively ineffective in separating the two groups (Kam & Meyer, 2015). Previous research has found that there are at least two types of careless respondent: one group chooses identical (or nearly identical) responses for consecutive items mindlessly; the second group chooses random responses throughout (Kam & Meyer, 2015; Maniaci & Rogge, 2014; Meade & Craig, 2012; see also Breitsohl & Steidelmuller, 2018). Presumably, these response patterns reflect a lack of motivation for carefully reading and responding to survey questions. In a typical survey, some items are related to each other in their meaning, while others are opposite to each other.

* Corresponding author at: Faculty of Education, The University of Macau, Macau.
E-mail address: ChesterKam@umac.mo (C.C.S. Kam).

https://doi.org/10.1016/j.paid.2018.03.022
Received 22 January 2018; Received in revised form 9 March 2018; Accepted 10 March 2018
Available online 20 March 2018

Table 1
Number (and percentage) of respondents passing each instructed response items and item-total correction.

Items	Number (and percentage) of respondents passing the item	Item-total correlation (correlation between an item and the entire scale excluding the item)
First item Skip this item, and do not choose any of the statements on this page, including this statement; ignore the instruction of choosing one of the (five) statements here. (This is one out of five options in a multiple-choice item. Before this item participants have been requested to read all options carefully before responding with the best option.)	726 (96.16%)	0.53
Second item Please select strongly disagree for this item.	705 (93.38%)	0.56
Third item Select strongly agree for this item.	713 (94.44%)	0.48
Fourth item I am competent in panabogy—skip this item to show that you have read survey questions carefully.	648 (85.83%)	0.70
Fifth item I like people in general and please skip this item to show you carefully read the questions.	580 (76.82%)	0.58

Therefore, it is unlikely that a motivated participant would choose identical or random responses to all these items. Instructed response items can thus identify participants who are not careful or motivated enough to follow instructions (Oppenheimer, Meyvis, & Davidenko, 2009).

However, other researchers have questioned the validity of instructed response items. Huang, Curran, Keeney, Poposki, and DeShon (2012) discovered that such items "flagged an unusually high rate of IER [insufficient effort respondents]" (p. 103). Therefore, these researchers did not include instructed response items in their analysis, and their results challenge the utility of instructed items. McGonagle et al. (2016) found that exclusion of careless respondents caused only a slight change in correlations. Curran and Hauser (2015) found that a small proportion of respondents failed to answer instructed response items correctly even if they were asked to read the items aloud. Finally, Niessen, Meijer, and Tendeiro (2016) used an instructed response item to identify careless respondents. They found that the method correctly flagged a large proportion of respondents instructed to respond carelessly, but also flagged a large proportion of respondents as careless even when instructed to respond carefully.

Thus, these results cast some doubt on the value of instructed response items as a measure of careless responding. If instructed response items exclude an *unreasonably* large number of respondents, the validity of this method as a whole may be called into question.

1.2. The current research

Instructed response items apparently have strong face validity or content validity—it is assumed that participants fail to answer these items due to carelessness. However, as noted, some researchers have questioned the method, and thus its validity needs to be reexamined. The purpose of the current investigation is to assess the validity of instructed response items.

The current study examines the convergent validity of instructed response items with a number of other potential indicators of careless responding: total survey completion time, longstring patterns, synonyms, antonyms, and odd-even consistency. Previous researchers have suggested that participants who are exceptionally quick to complete a survey are likely candidates of carelessness. Some careless respondents are also likely to select the same response option over multiple consecutive items without paying attention to the items, creating long strings of identical responses. Synonyms and antonyms could help identify carelessness, because careful respondents should choose similar responses to synonyms and contrary responses to antonyms. Moreover, a good index of careless responding should show discriminant validity with other response styles such as acquiescence (Kam & Meyer, 2015). Finally, Kam and Meyer (2015) showed that careless responding is likely to cause regular- and reverse-keyed items to correlate less strongly with each other. Presumably careless participants are less likely to attend to item wording, causing regular- and reverse-keyed items to correlate less well (particularly, less negatively) with one another. Therefore, we examined the consequences of excluding careless respondents using different cutoff scores for instructed response items. With stricter cutoff values, a valid index would show increasingly negative correlations between regular- and reverse-keyed items.

2. Method

The data were from 755 participants from the United States and Canada completing an online survey. Participants completed the survey in exchange for online bookstore coupons. The data have been published in a previous study (Kam, 2017), whose purpose was unrelated to the current one. The survey has approximately 300 questions, ranging from personality measurement (Big Five traits) to job satisfaction. The focus of the current investigation, however, is on various indices of careless responding. The indices were selected because they have been found to be good in detecting careless respondents (e.g., Mahalanobis distance; Meade & Craig, 2012), and because they have been recommended by methodologists (e.g., longstring; Curran, 2016; DeSimone, Harms, & DeSimone, 2015; Meade & Craig, 2012).

2.1. Careless responding indices

2.1.1. Instructed response items

Participants were warned about the existence of instructed response items and given an example of how to answer this type of item at the beginning of the survey. Five instructed response items were randomly dispersed throughout the survey. These items are listed in Table 1. The scores range from 0 to 5, with 5 being correct for all five. We will examine the internal consistency among the five items.

2.1.2. Mahalanobis distance

Mahalanobis distance is a popular measure of erratic, uncommon response patterns. It is often touted as a good measure of random careless responding (Meade & Craig, 2012). Participants whose response pattern deviates from the pattern of most others have a high value of Mahalanobis distance.

2.1.3. Longstring

Longstring (or, maximum longstring) is another popular index of careless responding. It represents the maximum number of times participants choose identical responses for consecutive questions. It thus