An item level evaluation of the Marlowe-Crowne Social Desirability Scale using item response theory on Icelandic Internet panel data and cognitive interviews

Vaka Vésteinsdóttir [a,*], Ulf-Dietrich Reips [b], Adam Joinson [c], Fanney Thorsdottir [d]

[a] University of Iceland, Department of Psychology, Aragata 14, 101 Reykjavik, Iceland
[b] University of Konstanz, Department of Psychology, Fach 31, 78457 Konstanz, Germany
[c] University of Bath, School of Management, Bath, UK
[d] University of Iceland, Department of Psychology, Sturlugata 3, 101 Reykjavík, Iceland

ARTICLE INFO

ABSTRACT

The Marlowe-Crowne Social Desirability Scale (MCSDS) is commonly used to validate other self-report measures within social and health research. Concerns over the scale's length (33 items) have repeatedly been raised. Nevertheless, prior efforts to develop psychometrically sound short forms of the MCSDS have not led to consistent findings. The purpose of this study was to develop a short form of the MCSDS, in accordance with guidelines for best practices in short form and scale development. Information on item properties, obtained with item response theory (IRT) and cognitive interviews (CogI), were used to eliminate items with poor properties and select items for a short form to be administered via the Internet. The IRT analyses were based on responses from 536 Internet panel members and the CogI sample consisted of 40 interviewees. Ten items were dropped due to poor psychometric properties and out of the 23 remaining items a ten item short form was developed.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The Marlowe-Crowne Social Desirability Scale (MCSDS) (Crowne & Marlowe, 1960) is a 33 item commonly used instrument for measuring social desirability response style (SDRS). SDRS is a respondent's tendency to present him- or herself favorably and can have confounding effects on self-reported data (see e.g. Kaufmann & Reips, 2008; Podsakoff et al., 2003; Podsakoff et al., 2012). It is therefore important, when using self-reports, to be able to obtain estimates of SDRS. However, adding a 33 item scale to a questionnaire places extra burden on respondents, potentially preventing the use of the MCSDS in research when cost and/ or respondent fatigue are of major concern. The aim of this study is therefore to extend the Vésteinsdóttir et al. (2015) study of the psychometric properties of the MCSDS, by evaluating single items from the MCSDS for the purpose of eliminating items with the weakest psychometric properties and suggesting a psychometrically sound short form of the scale.

1.1. Social desirability response style

Evidence consistent with SDRS comes from studies on self-reports of behaviors such as illicit substance use, alcohol use, smoking, abortion, energy consumption, income, criminal behavior, voting behavior, exercise and seat belt use (Tourangeau & Yan, 2007). SDRS is a serious problem in assessment because it can inflate the scores on desirable items and deflate the scores on undesirable items, resulting in biased estimates and possible distortion of relationships between variables. SDRS can therefore have a significant confounding effect on empirical findings and lead to misleading conclusions (see Cote & Buckley, 1988). Concerns about the effect of response biases on the validity of research findings have been shown to influence how reviewers perceive the quality of results and subsequent decisions about publication of work (Pace, 2010).

If the target variables are not related to a measure of SDRS, it can be concluded that they are free of SDRS and that their relationships are not distorted by the bias. If, however, SDRS is identified in the target variables, researchers must control for the effects of the bias in subsequent analysis (for an overview see for example Podsakoff et al., 2003). There are a number of statistical techniques for controlling for the effects of SDRS but the recommended one is to use a direct measure of SDRS in a latent variable model (Podsakoff et al., 2012).

---

* Corresponding author.
E-mail addresses: vakav@hi.is (V. Vésteinsdóttir), reips@uni-konstanz.de (U.-D. Reips), A.Joinson@bath.ac.uk (A. Joinson), fanneyt@hi.is (F. Thorsdottir).

1.2. The Marlowe-Crowne Social Desirability Scale

The MCSDS has been extensively used to validate target measures and control for the effect of SDRS. Between 1960 (when the MCSDS was first published) and 2002, >1000 articles and dissertations mentioned the use of the MCSDS when the PsychINFO, ERIC, Sociological Abstracts and Social Sciences Abstracts databases were searched (Beretvas et al., 2002) and according to Google Scholar in May 2016, >7000 works had cited the original article on MCSDS. The number of studies using the original MCSDS, or some of its short forms, continues to grow to this day (see van Schie et al. (2016) for a recent example of the use of the MCSDS, and Black and Reynolds (2016) for recent use of a MCSDS short form, in Internet administrated scale validation).

Some have, however, questioned the use of SDRS measures such as the MCSDS (see e.g. Tracey, 2016) due to a controversy over the scale's content and dimensionality (Barger, 2002; Fischer & Fick, 1993; Helmes & Holden, 2003; Leite and Beretvas, 2005; Loo & Thorpe, 2000; McCrae & Costa, 1983; Paulhus, 1984; Ventimiglia & MacDonald, 2012). Other researchers have opted for the BIDR (Paulhus, 1991) as a measure of SDRS, partly because of the often presumed two dimensional nature of SDRS (see e.g. Davis et al., 2012) and because the BIDR was developed with newer and more sophisticated techniques (Lambert et al., 2016). New research has, however shown that the MCSDS is unidimensional in an Internet administration (Vésteinsdóttir et al., 2015) and outperforms the BIDR in detecting faking (Lambert et al., 2016), suggesting both adequate psychometric properties and usefulness of the MCSDS.

A major limitation of the MCSDS is the length of the scale. The MCSDS consists of 33 true/false items, which describe behaviors that are "culturally sanctioned and approved of but which are improbable of occurrence" (Crowne & Marlowe, 1960, p. 350). Adding 33 items to an assessment procedure places extra burden on respondents and adds to the cost of administration. This is particularly true of instruments such as the MCSDS which are used for validation of other assessment tools and thus presented with at least one other measure. Increased length of a questionnaire can reduce potential participants' willingness to respond and increase the likelihood of exhausting respondents' patience (fatigue effect), which can result in non-completion of questionnaires and reduced response quality (Galesic & Bosnjak, 2009; Reips, 2010; Schuman & Presser, 1996). It would therefore be desirable to have a shorter measure of SDRS in order to have more space for questions on the assessment topic and reduce response burden and possible cost.

1.3. Short form development

For these reasons, researchers have attempted to develop short forms of the MCSDS scale, selecting items based on results from exploratory factor analysis (Ballard, 1992; Reynolds, 1982; Strahan & Gerbasi, 1972). Unfortunately, however, these attempts have not led to consistent findings (Vésteinsdóttir et al., 2015). Previous efforts have three main limitations: First, the short forms have been created using statistics which rely heavily on sample specific statistics. This is probably the main reason why previous attempts have not agreed on which items should be omitted from the short form. Secondly, the emphasis in previous studies was on selecting items for short forms to maximize internal consistency. However, focusing only on internal consistency in short form development may create a short form that is too narrow and potentially low in validity (Loevinger, 1954). Finally, short form developments have exclusively relied on convenience student samples (Vésteinsdóttir et al., 2015). Clearly, a student sample does not represent the population for which the scale is intended. The consequences of such sample non-representativeness can severely harm the short form development efforts (DeVellis, 2012). Furthermore, guidelines for best practices in scale and short form development recommend the use of multiple indicators of quality (Clark & Watson, 1995; DeVellis, 2012; Stanton et al., 2002).

One approach to overcome the limitations in previous short form development, listed above, is to use item response theory (IRT) to obtain information on item properties, instead of the previously employed techniques. The benefit of using IRT in short form development is that IRT models provide information on item properties in relation to respondents' estimated trait level (how much of the characteristic being measured, the respondent is presumed to possess). Taking respondents' trait level into account means that IRT estimates are not as highly dependent on the characteristics of the sample as methods that are purely based on item responses (see Embretson & Reise, 2000 for a more in depth explanation). The most commonly used short forms of the MCSDS have been developed using component factor analysis (Vésteinsdóttir et al., 2015), which is based on correlations between item responses. This method has the drawback of favoring redundant items (the more similar the items, the higher the correlation will be between them). With IRT the items are placed on a continuous scale, which represents the characteristic being measured. The items can thus be chosen to either measure as many points on the continuum as needed (e.g. when making a short form of a scale) or to have high precision at a certain point of the scale (see e.g. DeVellis, 2012; Embretson & Reise, 2000).

1.4. The present study

As the discussion above has highlighted, there is a need for a short and psychometrically sound version the MCSDS. In this study, items with the weakest psychometric properties will be identified and eliminated and a short form of the MCSDS (i.e. MCSD-SF) will be developed based on best practices in short form development. A combination of item response theory (IRT) and cognitive interviews (CogI) will be used to evaluate each item. In addition, previous factor analyses of the MCSDS will be included in the analysis to identify items that have repeatedly obtained the lowest factor loadings.

In view of the increasing number of studies that collect data online (e.g. Reips, 2012), the short form is intended for the Internet survey mode and thus the IRT analysis are based on Internet survey data with CogI, conducted face to face, for the purpose of evaluating the clarity of the items.

2. Method

2.1. Online survey

2.1.1. Participants and procedure

The IRT analysis was done on a sample collected by the Social Science Research Institute (SSRI) in July 2013 (see Vésteinsdóttir et al., 2015 for description). The survey, containing all 33 items of the MCSDS in Icelandic, was presented on three pages, each containing 11 items, to be consistent with other surveys sent out by the SSRI where lengthy question grids are generally avoided. An e-mail invitation was sent out to 1200 potential participants, drawn from the SSRI Internet panel. Duration of data collection was two weeks, with three reminders being sent out within the first 12 days after the original invitation was sent. Out of the 639 participants who responded to at least one item on the MCSDS, 536 participants (44.7% of the original sample) completed all items on the scale. Evaluation of psychometric properties of data obtained with the MCSDS was based on completed scales. The final sample consisted of 272 women (50.7%) and 264 men (49.3%), aged between 20 and 81 years (mean age being 49 years), with educational levels varying from elementary school education to a post-graduate university degree.

2.1.2. Instrument

The *Marlowe-Crowne Social Desirability Scale* (Crowne & Marlowe, 1960) contains 33 TRUE/FALSE items, 18 keyed in the true direction (attribution of desirable behavior) and 15 keyed in the reverse direction