



Acquiescence and social desirability as item response determinants: An IRT-based study with the Marlowe–Crowne and the EPQ Lie scales

Pere J. Ferrando *, Cristina Anguiano-Carrasco

Research Center for Behavioral Assessment, Rovira i Virgili University, Facultat de Psicologia, Carretera Valls s/n, 43007 Tarragona, Spain

ARTICLE INFO

Article history:

Received 7 July 2009

Received in revised form 9 December 2009

Accepted 14 December 2009

Available online 12 January 2010

Keywords:

Acquiescence

Social desirability

Marlowe–Crowne scale

Lie scale

Item factor analysis

Item response theory

ABSTRACT

The present study aims to shed some light on an old controversy about the joint impact of acquiescence and social desirability on item responses. There are two main hypotheses: (a) of the two biases, social desirability is by far the prime determinant, and (b) acquiescence operates in all sorts of items, including those impacted by social desirability. A new methodology is harnessed to assess these hypotheses in an empirical study based on two well known social desirability scales: the Marlowe–Crowne social desirability scale and the Lie scale of Eysenck's questionnaires. In both scales, the results suggest that even items which primarily measure social desirability can also be impacted to some extent by acquiescence.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Acquiescent responding (AR) and socially desirable responding (SDR) are considered to be the most important response biases in personality measurement (e.g. Paulhus, 1991). Furthermore, in real test taking situations, both types of bias are expected to operate simultaneously to a greater or lesser extent (e.g. Hofstee, Ten Berge, & Hendricks, 1998). However, the question of how AR and SDR jointly impact item responses has received surprisingly little attention in the vast literature on response biases. A review shows that two general positions have been considered. The first position assumes that, of both biases, SDR is by far the prime determinant. So, AR would mainly operate in “neutral” items, and its impact would be negligible in items that evoke the tendency to give socially desirable responses (Edwards, 1961; Jackson & Messick, 1961; Stricker, 1963). The second position assumes that AR operates in all sorts of items. So, items that are impacted by SDR can also be impacted by AR (Couch & Keniston, 1960; Diers, 1964).

To assess which of the positions above is the most correct, a relatively simple approach would be to study the impact of AR on responses to social desirability (SD) scales (or sets of items intended to measure solely SD). If the first position is correct, the impact of AR on items that are almost pure measures of SD should be negligible. However, if the second position is correct, some impact should be detected. And according to the most extreme formulation (Couch

& Keniston, 1960; Diers, 1964), this impact should be the same as is generally found in scales that measure substantive dimensions. This approach has been considered in a few studies (Bernhardson, 1970; Diers, 1964; Edwards, 1961; Jackson & Messick, 1961). However, these studies were mostly at the total scale level, and their results on the specific issue considered here were far from conclusive. Methodologically, the main problem encountered was how to clearly separate the impact of SDR from the additional impact of AR.

At present, probably the best procedure for controlling AR is to use a balanced set of items (Ray, 1983). In a well-designed balanced set, all the items are positively worded. However, half of the items measure in one direction of the dimension, whereas the other half measure in the opposite direction. In recent decades, factor analytic procedures that are based on balanced sets and which allow the impact of AR to be assessed in detail have been proposed (Billiet & McClendon, 2000; Ferrando, Lorenzo-Seva, & Chico, 2003). These procedures provide new tools for addressing this issue at the item level. However, they treat the item responses as continuous variables and require fully balanced scales, while most well-known SD measures use binary items and are only partially balanced (i.e. the number of positively keyed and negatively keyed items is different). Recently, Ferrando, Lorenzo-Seva, and Chico (2009) proposed a factor analysis (FA) based procedure that can be used with binary items and partially balanced scales and which is the one we shall use here.

The main purpose of the present study is to assess which of the two positions above is the most appropriate, and uses the direct approach based on analyzing well-known SD measures. The study

* Corresponding author. Tel.: +34 977558079; fax: +34 977558088.

E-mail address: perejoan.ferrando@urv.cat (P.J. Ferrando).

is solely concerned with the results obtained when these measures are administered in neutral conditions (i.e. voluntarily and by using standard instructions). It focuses on the item level (unlike previous studies), and uses the FA procedure proposed by Ferrando et al. (2009). It is hoped that this new analytical approach can provide clearer results on this issue.

The theoretical approach adopted here considers AR and SD as truly individual-difference variables that can be modelled as common factors (Eysenck & Eysenck, 1976; Ferrando et al., 2009) essentially independent from each other. Some authors (Hofstee et al., 1998; Stricker, 1963) have suggested that AR and SD are expected to be positively related, because they share a tendency to conformity or compliance. Empirical evidence, however, suggests they are essentially uncorrelated (Greenwald & Clausen, 1970; Stricker, 1963).

The two SD measures considered here are: the 21-item Lie scale which is included in the Eysenck Personality Questionnaires for adults (EPQ-A, Eysenck & Eysenck, 1975, and EPQ-R; Eysenck, Eysenck, & Barrett, 1985), and the 33-item Marlowe–Crowne social desirability (MCSD) scale (Crowne & Marlowe, 1960). Both use binary items and are partially balanced. Of the 21 Lie items, 15 describe undesirable behaviours and the remaining 6 describe desirable behaviours. Of the 33 MCSD items, 18 are socially desirable and 15 undesirable. Both scales are very popular, have received a great deal of attention, and have been intensively used in research. At the theoretical level, when administered under neutral conditions, both were intended to measure a general, unitary trait of SD mainly conceptualized as need for social approval (Crowne & Marlowe, 1960; Eysenck & Eysenck, 1975; Eysenck & Eysenck, 1976). This intention can be seen when the item contents are inspected. They mainly reflect approved behaviours that are believed to be unlikely to occur (desirably keyed items) or attitudes and practices that are socially undesirable but common, such as minor dishonesties, bad thoughts, weaknesses of character, etc. (undesirably keyed items).

Paulhus (1991) distinguished two components of SD within the general dimension. The first component, self-deceptive positivity, is conceptualized as an honest but overly positive view of oneself. The second component, Impression management, is more related to the traditional view of SD, and means that the respondent deliberately tailors his/her answers to create a more positive social image. According to Paulhus (1991), the Lie scales in Eysenck's questionnaires are almost pure measures of Impression management, whereas the MCSD scale tends to load more on the Impression management component, but is a less pure measure that also measures the first component to some extent. The procedures used here can also provide information on this issue.

Indeed, both scales have been factor analyzed in a large number of studies that aimed to assess their appropriate dimensionality. Reviews are provided in Ferrando, Chico, and Lorenzo (1997) and Lajunen and Scherler (1999) (Lie) and in Barger (2002) and Ferrando and Chico (2000) (MCSD). In this paper we shall not discuss these studies. Rather, we shall discuss various key points of how the scales were developed and analyze the characteristics of their items. This type of analysis has been considerably neglected in both scales.

The MCSD scale was developed by using judge agreement and item-total discriminatory power criteria (Crowne & Marlowe, 1960). Its item stem analysis suggests 'a priori' three types of problems. First, most item stems are long and complex. Second, 14-item stems are negatively worded, and the negative wording contains the modifiers "not" or "never". Furthermore, some of these items include double negatives (3 in the Spanish adaptation). Evidence shows that negatively worded items of this type tend to have poor measuring properties and might give rise to problems when assessing dimensionality and structure. Third, the content of three

pairs of items is highly redundant. Items 5 and 10 refer to lack of confidence in one's own abilities. Items 6 and 22 refer to doing things in one's own way. Finally, items 16 and 20 refer to admitting one's own errors. Pairs of redundant items such as these are likely to give rise to artifactual or semantic doublet factors (e.g. Kline, 1998).

The EPQ Lie scale was developed using FA procedures with the explicit aim of measuring a unitary factor (Eysenck & Eysenck, 1976, chap. 11). So, it is expected to have a clearer and stronger FA structure than the MCSD scale. Analysis of the item stems shows that they are all positively worded. They are also generally shorter and simpler, and refer to more specific behaviours than those of the MCSD scale.

1.1. The model and procedure

This section aims solely to provide a conceptual non-technical discussion that helps to interpret the present results. A more detailed methodological description can be found in Ferrando et al. (2009).

The general model used here is a bidimensional (content–acquiescence) item FA model intended for binary items, in which the item factor regressions are modelled as ogives. This model is a FA formulation of the bidimensional normal-ogive item response theory (IRT) model. For binary items, it is a more plausible model than the conventional FA model, which treats the direct binary scores as if they were continuous and unlimited variables. In practice, the main advantage of fitting this model instead of using conventional FA is that it partly avoids the attenuation problems created by the differential difficulties (i.e. proportions of endorsement) of the items. It (a) gives more correct estimates of the factor loadings, and (b) provides a more accurate assessment of the dimensionality of the data and of the appropriateness of the model.

The model has two stages: calibration and scoring. The calibration stage is a sequential FA in which (a) the acquiescence (ACQ) factor is first estimated, (b) its effect is partialled-out from the correlation matrix, and (c) the content factor is estimated from the corrected correlation matrix free from acquiescence. The procedure for fitting the model is minimum rank factor analysis (MRFA; Ten Berge & Kiers, 1991). MRFA is a least-squares FA procedure which has the additional advantage that it estimates the proportion of common variance accounted for by each of the two factors. These estimates are particularly relevant here for assessing the relative impact of the ACQ and SD factors.

The loadings on the first estimated ACQ factor are obtained by using a balanced subset of items, and they are all expected to be positive (the higher the acquiescence level, the greater the tendency to agree with all of the items). Next, the content factor is estimated from the "clean" correlation matrix, and it is expected to be bipolar: the positively keyed items are expected to have positive loadings, and the negatively keyed items, negative loadings.

We turn now to the scoring stage. In the model used here the factor scores are nonlinear transformations of the item scores, and their precision generally varies at different trait levels. However, as in standard FA, a single, marginal measure of reliability can be obtained for each factor. This measure can be used as an auxiliary index for assessing the relative importance of SDR and AR. Finally, the factor scores can be used for further validity assessment. Previous studies on the relations between the Lie and the SDS scales were based on correlations between the raw scores. Now, if AR also impacts the responses, these raw scores in fact reflect the mixed influence of both: SDR and AR. If this is so, assessment would be clearer if the relations between the SD factor scores on the one hand and the ACQ factor scores on the other were assessed separately.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات