Personality and Individual Differences 63 (2014) 36–40

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Measurement invariance tests of the Impression Management sub-scale of the Balanced Inventory of Desirable Responding

Brian K. Miller [a,*], Enrica N. Ruggs [b,1]

[a] Texas State University, Department of Management, 545 McCoy Hall, San Marcos, TX 78666, United States
[b] University of North Carolina at Charlotte, Department of Psychology, 9201 University City Blvd., Charlotte, NC 28223, United States

ARTICLE INFO

Article history:
Received 30 August 2013
Received in revised form 7 January 2014
Accepted 12 January 2014
Available online 15 February 2014

Keywords:
Measurement invariance
Multiple group analysis
Measurement equivalence
Impression Management
Anonymous
Confidential
Confirmatory factor analysis
Quasi-experiment

ABSTRACT

The Impression Management (IM) subscale of the Balanced Inventory of Desirable Responding (Paulhus, 1988) has been used as a proxy for common method variance in anonymous settings and as a cause of faking on personality testing in confidential settings. This study uses confirmatory factor analysis to conduct measurement invariance/equivalence tests (also known as multiple group analysis) on the IM subscale in a quasi-experiment in anonymous and confidential data collection settings. Using Brown's (2006) bottom-up approach to Cheung and Rensvold's (1999) model testing steps and Cheung and Rensvold's (2002) statistical tests, the IM sub-scale was determined to have equal form, equal factor loadings, equal indicator error variances, equal factor variance, and an equal factor covariance in both 1-factor and 2-factor models in both data collection settings. Mean scale scores were significantly higher in the confidential group than in the anonymous group. These results suggest that using the IM sub-scale as a cause of faking and as a proxy for common method variance (CMV) is likely to be acceptable because the psychometric properties of the instrument are invariant across testing conditions.

1. Introduction

Socially desirable responding (SDR) sometimes plays a role in how individuals respond to questionnaires, as they often will skew information about themselves (commonly referred to as faking) on non-cognitive measures (e.g., personality tests, non-cognitive diagnostic inventories) to present themselves more favorably (Donovan, Dwight, & Hurtz, 2003; Graham, McDaniel, Douglas, & Snell, 2002; McFarland & Ryan, 2006). Paulhus (1984) describes SDR as a two-dimensional construct comprised of self-deception (SD: a covert manipulation of how one perceives their self) and Impression Management (IM: an overt manipulation of how one is perceived by others). Impression Management in this context refers to one of two types of response bias; either a response set based upon testing conditions or internal motivation or a response style akin to a stable personality trait (Paulhus, 1991) and not to a collection of physical behaviors described by Bolino and Turnley (1999) based upon Jones and Pittman (1982) taxonomy that includes ingratiation and intimidation influence tactics. Regardless of whether IM is a response set or it is a response style, the extent to which IM arises may vary based on one's motivation to gain a particular incentive (Robie, Brown, & Beaty, 2007) and on one's opportunity or ability to fake (Tett, Freund, Christiansen, Fox, & Coaster, 2012). Such motivations and opportunities are often determined by whether the data are collected anonymously (and therefore responses cannot be tracked to respondents) or confidentially (in which case respondents are matched with their responses).

In the context of organizational surveys, researchers and practitioners are also concerned about the statistical influence that IM as a source of common method variance (CMV) can have on the predictor-criterion relationship (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) especially in cross-sectional single-source data collections (Moorman & Podsakoff, 1992). Remedies for CMV are described by Williams, Ford, and Nguyen (2004) such as including some measure of SDR and "using statistical means to partial out the associated methodological effects" (p. 383). Single-source cross-sectional research of this sort may result in "artifactual covariation...between constructs", (p. 885) (Podsakoff et al., 2003) because of the similar manner or source from which the data were collected. Podsakoff et al. (2003) suggest that both the characteristics of the items in a test as well as certain demand characteristics of the data collection mode can manifest itself as CMV and may affect the relationships between other measured variables. Concern for CMV dates back at least as far as Campbell and Fiske (1959)

* Corresponding author. Tel.: +1 512 245 7179; fax: +1 512 245 2850.
E-mail addresses: bkmiller@txstate.edu (B.K. Miller), eruggs@uncc.edu (E.N. Ruggs).
[1] Tel.: +1 704 687 1365.

http://dx.doi.org/10.1016/j.paid.2014.01.037
0191-8869/

and arguments both for and against the existence of CMV abound (see Richardson, Simmering, & Sturman, 2009 for a thorough review).

In anonymous settings researchers often statistically control for IM to account for CMV (Podsakoff et al., 2003), while in confidential settings, researchers often statistically control for IM to offset the impact of faking (e.g., personality inventories, employment selection tests) which are two decidedly different purposes. If appropriate for both purposes, then respondents' conceptualization of the scale items must be consistent across these purposes. It is therefore imperative to demonstrate that respondents in groups ascribe the same meaning to items in the scales (Cheung & Rensvold, 1999). It must be shown that the scale measures the same construct, in the same manner, across groups (Reise, Widaman, & Pugh, 1993). To address this issue, we conducted measurement invariance tests of an IM measure collected in both anonymous and confidential groups using confirmatory factor analysis (CFA). This invariance is essentially a test of the generalizability of the scale for different purposes, (e.g., as a statistical remedy for common method variance in anonymous settings and as an indicator of faking in confidential settings). Thus, the present research addresses two questions. First, do people conceive of a measure of IM in a similar manner in both anonymous and confidential setting? Second, if they do, are there true differences in scale means in these settings?

2. Methods

2.1. Sequence of invariance tests

Tests of equivalence or invariance are of two broad types: measurement level invariance and latent construct invariance (also known as structural invariance). Measurement invariance tests always precede structural invariance tests although the specific order of the tests within these two broad categories varies (Bollen, 1989; Byrne, 1998; Byrne, Shavelson, & Muthèn, 1989). This study used Brown's (2006) bottom-up approach and conducted the specific following tests: (1) tests of equal form, (2) tests of equal factor loadings, (3) tests of equal indicator error variances, (4) tests of latent factor variance, and (5) for the two-factor model a test of the covariance between latent factors. The first two are standard fare for invariance testing but the latter are considered stringent by some researchers (Byrne, 1998; Cheung & Rensvold, 1999; Kline, 2011). We did not conduct tests of scalar invariance which are essentially tests of whether item means differ between groups because to constrain item means to equivalence obviates one of the purposes of this research: to examine if mean scale-scores are meaningfully different in anonymous and confidential administrations.

2.2. Use of fit indices

Because the chi-square difference test is heavily influenced by sample size, Steenkamp and Baumgartner (1998) recommend that researchers examining instruments using measurement equivalence or invariance tests supplement the chi-square difference test with other fit indices. Cheung and Rensvold (2002) provide a rule of thumb regarding the comparative fit index (CFI; Bentler, 1990) and suggest that each subsequently more restrictive test should result in less than a −.01 decrement to fit in order to indicate invariance across groups. That is, if the CFI decreases in magnitude in a successively restrictive model by less than .01 (e.g., $\Delta CFI = -.009$) then the two models are essentially equivalent. It should be noted that there are no standard errors associated with this test, so it essentially serves as a guideline of sorts. While the original model fit in the first test used both the chi-square and the CFI as baselines, we also supplemented these baseline fit tests with the Root Mean Squared Error of Approximation (RMSEA; Browne & Cudeck, 1993) and the Standardized Root Mean Squared Residual (SRMR). Hu and Bentler (1999) recommend that good model fit is indicated when RMSEA <.06, CFI $\geqslant$.95, and SRMR <.08. However, strict rejection of models based upon rigid adherence to fit index cutoffs should be considered only with regard to theoretical or substantive issues (Hopwood & Donnellan, 2010; Marsh, Hau, & Wen, 2004).

2.3. Instrument

The current study used the Impression Management sub-scale of the Balanced Inventory of Desirable Responding (BIDR: Paulhus, 1988). This IM measure was chosen because it is one of the most widely used SDR measures (Steenkamp, de Jong, & Baumgartner, 2010). Two different models were examined in the analyses below: (1) a one-factor solution for the IM sub-scale and (2) a two-factor solution based on the results of Li and Li (2008) with all 10 reverse scored items forced to load solely on a denial factor and all 10 positively worded items forced solely onto an enhancement factor. The IM sub-scale (Paulhus, 1988) was administered using a one-to-seven Likert response scale anchored by "strongly disagree" and "strongly agree" and scored continuously. An example of an enhancement item is "I always obey laws even if I'm unlikely to get caught" and a denial item is "When I was young I sometime stole things" (reverse scored). In both the anonymous and confidential conditions, the one-factor model resulted in Cronbach's coefficient alpha of internal consistency reliability for scores of .78. In the anonymous condition alpha was .72 for the 10-item denial factor and .61 for the 10-item enhancement factor. In the confidential condition alpha was .70 for the denial factor and .65 for the enhancement factor.

2.4. Data collection

Data were collected using a quasi-experimental framework (Campbell & Stanley, 1963; Cook & Campbell, 1979) in two large sections of a required undergraduate business class at a mid-sized public university in the southwestern U.S. A coin flip decided that one class would respond to the questionnaire anonymously and therefore not affix their name to their survey and the other class would respond confidentially and therefore respondents were required to put their name on their survey. Participants in the anonymous setting were told that this questionnaire was part of an attempt to develop a new measure of personality but there would be no way to match respondents to responses. Those in the confidential setting were told that this instrument would allow their professor to understand their individual personalities so it was imperative that the professor be allowed to match respondents to responses. Complete data were collected from 436 respondents.

2.5. Participants

2.5.1. Anonymous group

Two hundred sixteen respondents anonymously completed the IM sub-scale of the BIDR in one section of the course. Of those, 123 were male (56.9%), 92 were female (42.6%), and one did not indicate their gender. The mean age was 22.38. The self-reported racial/ethnic group membership was: 71.6% White, 6% Black, 16.3% Hispanic, 3.7% Asian, 0.5% American Indian, and 1.9% other.

2.5.2. Confidential group

Two hundred twenty respondents confidentially completed the IM instrument in the other section of the course. Of those, 116 were male (53%) and 103 were female (47%). The mean age was 22.16. The self-reported racial/ethnic group membership was: