Journal of Research in Personality 45 (2011) 417–429

Contents lists available at ScienceDirect

Journal of Research in Personality

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

The ubiquity of common method variance: The case of the Big Five

Michael D. Biderman [a,*], Nhung T. Nguyen [b], Christopher J.L. Cunningham [a], Nima Ghorbani [c]

[a] Department of Psychology, The University of Tennessee at Chattanooga, United States
[b] Department of Management, Towson University, United States
[c] Department of Psychology, University of Tehran, Iran

ARTICLE INFO

Article history:
Available online 19 May 2011

Keywords:
Big Five structure
Bifactor models
Common method variance
Confirmatory factor analysis
Method bias
Multitrait–multimethod
Personality

ABSTRACT

The factor structures of the International Personality Item Pool (IPIP) and NEO-FFI Big Five questionnaires were examined via confirmatory factor analyses. Analyses of IPIP data for five samples and NEO data for one sample showed that a CFA model with three method bias factors, one influencing all items, one influencing negatively worded items, and one influencing positively worded items fit the data significantly better than models without method factors or models with only one method factor . With the method factors estimated, our results indicated that the Big Five dimensions may be more nearly orthogonal than previously demonstrated. Implications of the presence of method variance in Big Five scales are discussed.

1. Introduction

In the past 30 years there has been resurgence in the study of personality in psychology due primarily to the discovery of a common factor structure underlying measures of personality characteristics. The dominant taxonomy is a lexically based five-factor structure originally developed within countries that use Northern European languages (e.g., Saucier & Goldberg, 2003). Most popularly known as the Big Five, this framework includes the traits of Extraversion (E), Agreeableness (A), Conscientiousness (C), Emotional Stability (S, often measured as Neuroticism), and Openness to Experience (O, sometimes measured as Intellect).

Despite the wide acceptance and application of this personality framework, several measurement-related issues have continued to challenge personality researchers. In particular, although conceived of as orthogonal dimensions of personality, correlations between summated scale scores on most Big Five personality tests are generally moderately positive (e.g., Digman, 1997; Mount, Barrick, Scullen, & Rounds, 2005). There are at least two explanations for this. The first is that the five factors commonly estimated are actually themselves correlated and perhaps indicators of higher order factors. More specifically, it has been suggested that the Big Five factors are indicators of the higher order factors of *stability* (as indicated by agreeableness, conscientiousness, and the inverse of neuroticism) and *plasticity* (as indicated by openness and extraversion) (DeYoung, Peterson, & Higgins, 2001; Digman, 1997). Others have alternatively suggested that there may be one overriding personality factor, deemed *evaluation* (Goldberg & Somer, 2000; Saucier, 1997), the "Big One" (Musek, 2007) or the general factor of personality (GFP) (Rushton, Bons, & Hur, 2008; Van der Linden, Nijenhuis, & Bakker, 2010).

A second explanation for the commonly identified positive relationships among Big Five scale scores is that there is a separate source of influence that affects responses to all items in these questionnaires, and that this influence is somehow distinct from that of the Big Five factors themselves. Often this type of shared influence across scores collected using a specific method is referred to as common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The word bias in this context refers to the individual differences that become manifest when the same method is used across multiple personality scales. Associated with this common bias is the notion of common method variance, which, in the present context, can be understood as variance in Big Five scale item responses throughout a measure that is due to the influence of common method bias.

The existence of common method variance has been recognized in questionnaire research for many years (e.g., Cote & Buckley, 1987; Doty & Glick, 1998). Most research on this topic has been based on analyses of multitrait–multimethod data from a single measure, usually an isolated scale or domain score, per trait-method combination. Although helpful in highlighting the potential impact of common method bias, such investigations have not made it possible to separate within-dimension covariance from between-dimension covariance (Tomas, Hontangas, & Oliver, 2000). Indeed, until recently, the study of common method variance based on

* Corresponding author. Address: Department of Psychology, The University of Tennessee at Chattanooga, 615 McCallie Avenue, Chattanooga, TN 37403, United States.
E-mail address: Michael-Biderman@utc.edu (M.D. Biderman).

0092-6566/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jrp.2011.05.001

analyses of individual items or representative item parcels has been neglected. This is unfortunate, given that such analyses are necessary to properly estimate and compare within- and between-dimension variability (Marsh, Scalas, & Nagengast, 2010).

One of the first studies to permit this type of variability separation using Big Five measure data was Schmit and Ryan (1993), in which analyses of multiple item composites from each dimension revealed the potential for measures of the Big Five traits to include common method variance. Schmit and Ryan factor analyzed responses to item composites of the NEO-FFI (Costa & McCrae, 1989) within a work context using applicant and non-applicant samples. An exploratory factor analysis (EFA) of the non-applicant sample demonstrated the expected five-factor solution, but in the applicant sample, a six-factor solution fit the data best. Schmit and Ryan labeled this sixth factor an "ideal employee" factor, noting that it, "included a conglomerate of item composites from across four of the five subscales of the NEO-FFI" (Schmit & Ryan, 1993, p. 971). Interestingly, items from all five NEO-FFI subscales loaded on this factor, suggesting that the "ideal employee factor" represented a form of common method bias.

Beginning in the late 1990s confirmatory factor analyses (CFAs) were conducted of questionnaire items or parcels to identify and study method biases. These studies included analyses of data collected with the Rosenberg Self-Esteem scale (Marsh, 1996; Marsh et al., 2010; Motl & DeStefano, 2002; Tomás & Oliver, 1999) and investigations of the possibility that common method bias may represent or reflect respondent faking or socially desirable responding in certain situations (e.g., Biderman & Nguyen, 2004; Bäckström, 2007; Bäckström, Björklund, & Larsson, 2009; Cellar, Miller, Doverspike, & Klawsky, 1996; Klehe et al., 2011; Ziegler & Buehner, 2009). In these experimental studies, participants were typically asked to respond to Big Five measure items under faking and no-faking conditions. In the faking conditions of these studies, variance common to all items was represented by a single latent variable similar to what Podsakoff et al. (2003) labeled an "unmeasured method" effect. However, with the exception of Bäckström (2007) and Bäckström et al. (2009) the study of common method variance in Big Five questionnaires in nonapplicant honest conditions has received little attention. This is problematic, given that in most scenarios, participants are instructed to do precisely that – respond honestly.

This limitation of previous research in this area combined with the fact that most personality assessment is by self-reported completion of personality inventories, leaves a major deficit in our understanding of what is actually being assessed when we use common personality measures such as those designed to capture the Big Five traits. Further complicating matters is a common recommendation for developing and/or choosing assessment items for self-report measures, that encourages the inclusion of both positively worded items ("I am the life of the party") and negatively worded items (e.g., "I don't talk a lot") in a single scale. The logic behind including both types of items is that their presence might reduce the effects of participant response tendencies such as acquiescence (DeVellis, 1991; Nunnally, 1978; Spector, 1998). This recommendation has been so widely shared that the practice of using negatively worded items to presumably counteract respondents' acquiescence can be found throughout most areas of organizational research including personality assessment (e.g., Paulhus, 1991), leadership behavior (e.g., Schriesheim & Eisenbach, 1995; Schriesheim & Hill, 1981), role stress (Rizzo, House, & Lirtzman, 1970), job characteristics (Harvey, Billings, & Nilan, 1985), and organizational commitment (e.g., Meyer & Allen, 1984).

Unfortunately, the negatively worded items that were introduced to counter individuals' response tendencies have been found to increase systematic and perhaps construct-irrelevant variance in scale scores in studies: (a) of self-esteem (e.g., Hensley & Roberts, 1976; Marsh, 1996; Marsh et al., 2010; Motl & DeStefano, 2002; Tomás & Oliver, 1999), (b) using Rizzo, House, and Lirtzman's (1970) role conflict and role ambiguity scale (McGee, Ferguson, & Seers, 1989), (c) using Meyer and Allen's (1984) Organizational Commitment scale (Magazine, Williams, & Williams, 1996), (d) using Spector's (1988) Work Locus of Control Scale, and (e) using Hackman and Oldham's (1975) Job Diagnostic Survey (Idaszak & Drasgow, 1987). In addition to increased "noise" interjected by such items and the potential multidimensionality introduced by negatively worded items, the inclusion of such items in leadership behavior measures has been shown to decrease a scale's reliability and validity (Schriesheim & Eisenbach, 1995; Schriesheim & Hill, 1981).

Recently, Marsh et al. (2010), using confirmatory factor analyses, provided evidence for two conclusions regarding the factorial structure of questionnaires employing negatively worded items. First, Marsh et al. found that a model with two method factors (one influencing only positively worded items and the other influencing only negatively worded items) fit the data of the Rosenberg Self Esteem (RSE) scale better than models without method factors and better than models with only one wording-type factor. Although other researchers had found item wording influences associated with negatively worded items (e.g., DiStefano & Motl, 2006), Marsh et al.'s results provided evidence for analogous influences associated with positively worded items. Marsh et al. further found, based on longitudinal models, that positive and negative influences were not sporadic and spontaneous, but substantive and stable over time. These two findings coupled with other studies in which method factors have been implicated (Cote & Buckley, 1987; Doty & Glick, 1998) suggest that method effects including item-wording specific method effects may be influential whenever personality is assessed using self-report questionnaires.

Given the mounting evidence for the prevalence of common method variance in personality assessment and the increasing usage of personality assessments in organizational research and practice, it is surprising that few attempts have been made to examine the effects of method bias and item wording biases on the factor structure of Big Five measures. As mentioned previously, studies estimating a common method factor have for the most part focused on identifying socially desirable responding. Apart from those above-mentioned studies, there have been no published CFA models of Big Five questionnaire data that have included item-wording factors. For all the reasons already stated, the main purpose of the present study was to closely examine the factor structures of two commonly used Big Five questionnaires, the IPIP and NEO-FFI with the intent of assessing the extent to which responses to items in these questionnaires are influenced by method factors and/or wording-specific method factors. This was done in a fashion similar to that used by Marsh et al. (2010), by comparing CFA models with different assumptions concerning general method factors and wording-specific method factors.

The specific models that were compared in this study are presented in Fig. 1. Within this figure, Model 1 is a basic CFA of a Big Five questionnaire with correlated trait factors but no method factor. If there were a common method influence on all 50 items of this instrument, the presence of such an influence would have to be accommodated in the model by increased positive correlations among the factors (Paglis & William, 1996; Williams & Brown, 1994). In Model 2, a single method factor, M, has been added to the basic CFA of Model 1 (e.g., Bäckström, 2007; Bäckström et al., 2009; Cellar et al., 1996). In this model, M is defined as an "unmeasured" method factor in that it has no unique indicators but rather is estimated from indicators of the Big Five factors. M is a first order factor whose indicators are items which also are indicators of the Big Five factors, not a higher order factor. This type of model