An empirical investigation of the influence of social desirability on the factor structure of the Chinese 16PF

Jinyan Fan *, Corbin C. Wong, Sarah A. Carroll, Felix J. Lopez

Department of Psychology, Hofstra University, 100 California Avenue, Hempstead, NY 11549, United States

ABSTRACT

This study examined whether social desirability affects the factor structure of the Chinese 16PF. Participants were 4645 Chinese first-year undergraduate and graduate students who completed the Chinese 16PF as part of a mandatory psychological assessment. Based on students' score on the Impression Management scale (Paulhus, 1998), two groups were formed: (a) the high socially desirable responding group, consisting of students whose IM scores fell within the upper 20% of the distribution, and (b) the low socially desirable responding group, consisting of students whose IM scores fell within the bottom 25% of the distribution. A series of exploratory and confirmatory factor analyses indicated that the factor pattern and factor loadings of the Chinese 16PF were invariant across these two groups.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

During the last two decades, personality measures have been widely used as an employee-selection instrument, largely due to accumulating meta-analytic evidence that personality scores are predictive of a number of important organizational outcomes such as job performance (Barrick & Mount, 1991) and employee satisfaction (Judge & Bono, 2001). Furthermore, research has illustrated that personality measures typically do not display group mean-score differences large enough to result in adverse impact (e.g., Foldes, Duehr, & Ones, 2008).

Despite these benefits, many organizations are concerned that in organizational settings, particularly in selection contexts, individuals are motivated to elevate their personality scores in order to obtain beneficial outcomes (Rosse, Stecher, Miller, & Levin, 1998). This score elevation, or the tendency to present an overly positive self rather than the true self, has been generally referred to as socially desirable responding (SDR) (Douglas, McDaniel, & Snell, 1996). In the literature, SDR is either evoked by faking good instructions in experimental studies, assumed in high motivation testing contexts, or measured by various social desirability inventories.

Research has demonstrated that individuals *can* engage in SDR if asked to (e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Viswesvaran & Ones, 1999), and that some applicants *do* engage in SDR in selection contexts (e.g., Anderson, Warner, & Spencer, 1984). Furthermore, SDR does not appear to affect personality

measures' criterion-related validity in real selection contexts (e.g., Hough et al., 1990; Ones, Viswesvaran, & Reiss, 1996). Despite the above consensuses, scholars still disagree on two fundamental issues.

The first issue concerns the base rate of SDR among job applicants. Using innovative techniques, several researchers have estimated that the base rate of mild SDR ranges from 25% to 50%, while that of extreme SDR ranges from 5% to 10% (e.g., Donovan, Dwight, & Hurtz, 2003; Griffith, Chmielowski, & Yoshita, 2007). Nevertheless, two recent studies (i.e., Ellingson, Sackett, & Connelly, 2007; Hogan, Barrett, & Hogan, 2007) based on large sample sizes provided contrary evidence. Ellingson et al., using a within-subject design, investigated whether the same employees tended to score higher on the California Personality Inventory (CPI) within a selection context than within a developmental context. Rather small mean-score differences across the two contexts were found, with an averaged effect size of only .075. Hogan et al. examined whether applicants who had failed a pre-employment screening, which included the Hogan Personality Inventory (HPI), an ability test, and a language proficiency test, would be able to elevate their HPI scores when reapplying for the same position six months later. These authors reported minimum mean-score elevation over retesting.

The second point of debate concerns whether SDR negatively affects the construct validity of personality tests. Among the many aspects of construct validity, factor structure or factorial validity has received the most research attention in this literature. If SDR compromises the factor structure of a personality test, then the personality test no longer measures the intended traits, and any apparent inference that the personality test will predict job

---

* Corresponding author. Tel.: +1 516 463 6349.
E-mail addresses: Jinyan.Fan@hofstra.edu, fanjinyan@yahoo.com (J. Fan).

performance cannot be justified (Binning & Barrett, 1989). In the following section, we review previous research on SDR's effect on personality factor structure and then explain how the present study extends this line of investigation.

To date, at least three research strategies have been implemented to investigate SDR's influence on the factor structure of personality measures. The first strategy involves examining the effect of "directed faking" instructions using experimental methodologies (e.g., Douglas et al., 1996; Ellingson, Sackett, & Hough, 1999). For example, Ellingson et al. directed a sample of enlisted army personnel to complete the Assessment of Background and Life Experience (ABLE) two times: once under honest instructions and once under "faking good" instructions. These authors found that while the ABLE's supposedly multi-factor structure was retained in the honest condition, it deteriorated into a uni-factor structure in the "faking good" condition. A major limitation of this research strategy, however, is that it only shows responses elicited by "faking good" instructions can alter personality factor structure. It is by no means proven that individuals in lab studies display the same type of response behavior as individuals in real selection contexts (Smith, Hanges, & Dickson, 2001).

To address the above mentioned design-related weakness, the second research strategy allows SDR to occur naturally and examines the comparability of personality factor structure between job applicants and non-applicants (e.g. job incumbents and students) (e.g., Montag & Comrey, 1990; Schmit & Ryan, 1993; Smith & Ellingson, 2002; Smith et al., 2001). In one such study, Schmit and Ryan (1993) found that the Five Factor Model was upheld in a student sample, but not in an applicant sample. In the applicant sample, a sixth factor emerged that loaded highly on the facets of four of the Big Five factors. Schmit and Ryan labeled this factor "the ideal-employee factor" and interpreted its presence as an indication of applicant SDR. Montag and Comrey (1990) also found a similar sixth factor in a Big Five personality measure within an applicant sample.

Conversely, two recent studies based on several thousand participants (i.e., Smith & Ellingson, 2002; Smith et al., 2001) failed to replicate these prior findings, documenting instead strong evidence for factor structure invariance between applicant and non-applicant samples on the Reduced HPI (HPI-R; Smith et al., 2001). Based on a meta-analysis, Bradley and Hauenstein (2006) found that sample type (applicant vs. incumbent) had very small moderating effects on the inter-correlations of Big Five factors as measured by the NEO-FFI and HPI. Furthermore, this small moderating effect did not result in significant differences in the second-order factor pattern and factor loadings for these Big Five factors across different sample types.

It should be noted that positive findings of factor structure difference were based on modest samples, whereas negative findings were based on large samples. Because large sample studies are less prone to sampling errors than modest sample studies, one may argue that more credence should be given to the results of large sample studies by Smith and colleagues and Bradley and Hauenstein. Ellingson, Smith, and Sackett (2001) posited another explanation for these mixed findings. They suggested that there is a potential weakness inherent in comparing applicant and non-applicant samples because this strategy assumes applicants engage in SDR due to a strong situational demand, whereas non-applicants do not because there is little motivation to do so. However, as discussed earlier, research suggests that a high percentage of job applicants (i.e., 50–80%) do not engage in SDR. As a result, the potentially negative effect of SDR on personality factor structure might have been masked by the presence of a majority of relatively honest respondents.

Thus, Ellingson et al. (2001) concluded that this strategy does not *directly* address whether SDR damages the factor structure of personality measures and suggested a new, two-stage research strategy, which we refer to as the third research strategy. In stage 1 of the third research strategy, two groups are selected from the entire sample: one group consisting of respondents with the highest SDR scores and the other consisting of respondents with the lowest SDR scores. Stage 2 entails testing whether personality factor structure is invariant across the high vs. low SDR groups.

Ellingson et al. (2001) tested this strategy in four separate large datasets involving four widely-used personality inventories: ABLE, CPI, 16PF, and HPI-R. The motivational context varied across these four datasets, with some consisting of job applicants only, while others consisted of combinations of applicants, incumbents, and students. The criteria used for comparison group selection also varied, with the cutoffs for high-SDR groups ranging from the upper 10–33% of various social desirability score distributions, and the cutoffs for low-SDR groups ranging from the bottom 15–33% of the distributions. Results indicated that across these four personality measures, the high- and low-SDR groups had the same factor loading pattern and equivalent factor loadings. As such, Ellingson et al. concluded that SDR had little influence on the factor structure of personality measures.

The purpose of the present study was to replicate and extend Ellingson et al.'s (2001) research in the following ways. First, and most importantly, according to Paulhus' (1984) model, social desirability has two components: (a) self-deceptive enhancement, which represents unconscious favorability bias, and (b) impression management, which refers to intentional response distortion. Whereas the former reflects some substantive personality traits, the latter is sensitive to situational demands (Paulhus, 1998). Impression management is the component that is of the greatest concern to researchers and practitioners. Unfortunately, none of the four personality inventories in Ellingson et al.'s study included a social desirability scale that specifically measured the impression management component. As a result, we do not know if and the extent to which high SDR respondents in Ellingson et al.'s study engaged in intentional impression management. To address this issue, we used the Impression Management (IM) scale developed by Paulhus (1998).

Second, given that the vast majority of social desirability research has been conducted in Western societies, research on SDR in other cultures is needed. In particular, it is unknown whether Western findings regarding the effects of SDR on personality factor structure may generalize to non-Western societies. Filling this gap, we applied Ellingson et al.'s (2001) strategy within a sample of Chinese individuals who responded to the Chinese version of the 16PF in a context where the motivation to engage in SDR was high.

Third, previous large sample social desirability research (e.g., Bradley & Hauenstein, 2006; Ellingson et al., 2001; Smith & Ellingson, 2002; Smith et al., 2001) essentially combined numerous small sample studies. The different testing contexts of the participants in these studies is an extraneous variable that may have confounded past results. In contrast, participants in the present study represented a *single* large sample that shared exactly the same testing experience, thus minimizing these confounding concerns.

2. Method

2.1. Sample and procedure

Participants were first-year undergraduate and graduate students enrolled at a large east-coast university in the People's Republic of China. At this university, all incoming students are mandated to complete a psychological assessment (coordinated by the Student Counseling Center), consisting of the Chinese versions of the 16PF and the Minnesota Multiphasic Personality Inventory (MMPI). We randomly embedded 20 IM items (Paulhus,