Journal of Research in Personality 44 (2010) 53–61

Contents lists available at ScienceDirect

Journal of Research in Personality

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

The measurement equivalence of Big-Five factor markers for persons with different levels of education

Beatrice Rammstedt [a,*], Lewis R. Goldberg [b], Ingwer Borg [a]

[a] Leibniz Institute for the Social Sciences, GESIS, Mannheim, Germany
[b] Oregon Research Institute, Eugene, OR, United States

ARTICLE INFO

Article history:
Available online 24 October 2009

Keywords:
Big Five
Measurement equivalence
Acquiescence response bias
Factor structure

ABSTRACT

Previous findings suggest that the Big-Five factor structure is not guaranteed in samples with lower educational levels. The present study investigates the Big-Five factor structure in two large samples representative of the German adult population. In both samples, the Big-Five factor structure emerged only in a blurry way at lower educational levels, whereas for highly educated persons it emerged with textbook-like clarity. Because well-educated persons are most comparable to the usual subjects of psychological research, it might be asked if the Big Five are limited to such persons. Our data contradict this conclusion. There are strong individual differences in acquiescence response tendencies among less highly educated persons. After controlling for this bias the Big-Five model holds at all educational levels.

1. Introduction

In personality psychology, most research is based on samples of college students, especially undergraduates taking a course in psychology. This narrow subject population leads to the question of the extent to which the findings from such studies can be generalized to a broader population. Expressed in a more formal way: Are multi-item personality inventories transportable in the sense of measurement equivalence over subsamples of the population? If measurement equivalence (Drasgow & Probst, 2004; Vandenberg & Lance, 2000) does not hold for specific subsamples, it becomes dubious to compare the scale scores of persons from these different subsamples, because their scores are indicators of constructs that differ in terms of their structure and, thus, ultimately also in their meaning.

Over the last two decades, the Big-Five factors have become the most prominent model for describing the structure of personality traits. The Big Five have been found to replicate across different methods of data collection such as self-ratings, peer ratings, and behavioral descriptions, as well as across at least some different languages and cultures. However, personality research is still largely based on the typical "psych 100" student samples, which suggests that test norms may be inappropriate for other samples of individuals. Moreover, even the assumed five-factor structure may not hold in general. Indeed, several researchers investigating the Big Five in samples other than undergraduate populations using standardized and well-established inventories found that the five-factor structure did not replicate in their data sets (e.g., Körner, Geyer, & Brähler, 2002; Lang, Lüdtke, & Asendorpf, 2001; Mõttus, Allik, & Realo, 2007; Rolland, Parker, & Stumpf, 1998; Tokar, Fischer, Snell, & Harik-Williams, 1999; Toomela, 2003a). Table 1 provides an overview of the samples investigated and the instruments assessed in these studies, as well as the findings. As can be seen from the table, a variety of different well-established Big Five questionnaires such as the NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992), the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1989), or the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991; John & Srivastava, 1999) were used as measures of the Big Five in these studies. Moreover, because results not supporting the assumed model are less likely to be published, the studies presented in Table 1 could be just the tip of an iceberg of similar findings. In the present study, we investigate the structure of a Big-Five factor marker in subsamples that vary in their educational level. If the five-factor structure does not replicate in population-representative samples, it may nevertheless still hold in subsamples of persons with a higher educational background.

Specifically, in the present study we investigate whether the five-factor structure replicates in other than student populations, namely in samples that are representative of the general population. Therefore, we analyze the factor structure of a Big Five measure in two large data sets, both of which are representative of the general German adult population, allowing us to cross-validate our results simultaneously.

1.1. Individual differences in response bias

Individuals can differ in their mean response across all items, in the dispersion of their responses around their personal means,

* Corresponding author. Address: GESIS – Leibniz Institute for the Social Sciences, PO Box 12 21 55, D-68072 Mannheim, Germany. Fax: +49 6211246100.
E-mail address: beatrice.rammstedt@gesis.org (B. Rammstedt).

0092-6566/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jrp.2009.10.005

Table 1
Overview of studies failing to replicate the Big Five structure in non-student samples.

Authors	Year	Big-Five instrument	Sample	Results
Körner, Geyer, and Brähler	2002	NEO-FFI	$N = 1908$; population representative	Poor fit of varimax solution
Lang, Lüdtke, and Asendorpf	2001	BFI	$N = 480$; quoted with regard to age	Poor fit of varimax solution
Mõttus, Allik, and Realo	2007	NEO-FFI	$N = 1342$; population representative	Differences in explained variance and in congruence with simple structure among five educational groups
Rolland, Parker, and Stumpf	1998	NEO-PI-R, NEO-FFI	$N = 447$ college students and $N = 268$ military recruits	Weaker fit of the varimax solution in the military sample for both NEO-PI-R and NEO-FFI
Tokar, Fischer, Snell, and Harik-Williams	1999	NEO-FFI	$N = 485$ non-student adults	Poor fit of the varimax solution
Toomela	2003	NEO-PI-R	$N = 912$ male military members	Weaker fit of varimax solution than in the normative sample

and/or in the shape of their response distributions as reflected in measures of skewness and kurtosis. The results of several recent studies suggest that the psychometric quality of questionnaires decreases in general with lower levels of intelligence, education, and age (e.g., Allik, Laidra, Realo, & Pullmann, 2004; Mõttus, Allik, & Pullmann, 2007; Soto, John, Gosling, & Potter, 2008). This effect of lower psychometric quality seems to be associated with individual differences in subjects' use of the response options (cf. Gudjonsson, 1986; Mõttus, Allik, & Pullmann, 2007; Sigelman, Budd, Spanhel, & Schoenrock, 1981; Soto et al., 2008; Tokar et al., 1999), in particular by individual differences in acquiescent responding. Acquiescence refers to the tendency of an individual to consistently agree to questionnaire items, regardless of the content of the items (Jackson & Messick, 1958; Javeline, 1999). The tendency for acquiescent responding can be investigated by comparing the individual's responses to true- and false-keyed items assessing the same construct (e.g., Prefer to be with others, and Like to be all by oneself). When there are the same number of true- and false-keyed items per dimension, persons with a high tendency for acquiescence will have comparatively higher means scores than those with a lower such tendency.

Results of previous studies have shown, that such "yea-saying" appears to be more frequent among persons with lower social status, lower educational levels, and/or lower intelligence (e.g., Ayidiya & McClendon, 1990; Gove & Geerken, 1977; Krosnick, Narayan, & Smith, 1996; Lenski & Leggett, 1960; Narayan & Krosnick, 1996). It has been suggested that persons with relatively low education have less clear self-concepts, smaller vocabularies, and less verbal comprehension skills than more highly educated persons. This may make them relatively uncertain when it comes to responding to questionnaire items and thus leaves more room for the influence of systematic response biases (i.e., Goldberg, 1963). In the extreme, a respondent's answers could be completely independent of the content of the particular item. Even more moderate tendencies to agree with all items provides an additional source of variance, specific to each individual, that damages the psychometric properties of the scales and blurs the correlational patterns among them.

1.2. Controlling for response bias

One method to control for individual differences in acquiescence response tendencies is to compute each individual's mean across all items, subtract that mean from each response, and then analyze these "deviation" scores.[1] An alternative and more common procedure is to "ipsatize" the responses of each individual by using standard (z) scores (dividing an individual's deviation scores by the standard deviation of those scores), and thus simultaneously controlling for individual differences in both means and dispersions. With ipsatized data, every participant has the same mean and standard deviation across the total item pool, although individual differences in profile shape remain. In the present study, we compared both methods of controlling for individual differences in response bias.

We hypothesize that the usual Big-Five factor structure may not hold well in samples that represent the population of an entire country and, in particular, it is expected not to hold for samples of persons with lower educational levels. Moreover, we hypothesize that lower educated persons will have a stronger tendency for acquiescence response bias. Finally, when statistically controlling for acquiescence, we hypothesize that the Big Five structure can be replicated in all educational subsamples.

2. Methods

2.1. Samples and procedure

Analyses are based on the German surveys of the International Social Survey Program (ISSP) conducted in 2003/2004 and in 2005/2006, respectively. The ISSP household surveys were conducted together with the 2004 and 2006 versions of the German General Social Survey (ALLBUS). The ALLBUS was administered as a 45-min face-to-face interview, and the ISSP was a drop-off questionnaire to be completed by the respondents on their own after the personal interviews. The interviewer remained in attendance and collected the completed forms.

The ISSP survey was based on a representative sample of German adults (age $\geqslant 18$) living in private households in Germany. Foreigners residing in Germany and able to complete the questionnaire in German were included. The sample was drawn in a two-stage design from official registers of inhabitants kept by municipalities throughout Germany. First, the communities and sample points were selected randomly, and then individuals were randomly selected from each sampling point. Full details of the sampling are presented in the methods report on ALLBUS 2004 and ALLBUS 2006, respectively (Haarmann, Scholz, Wasmer, Blohm, & Harkness, 2006; Wasmer, Scholz, & Blohm, 2007).

Sampling and data collection were conducted by a commercial vendor (TNS-Infratest). Participation in the study was voluntary and not financially rewarded. The response rate was 41% both in 2004 and in 2006. Compared to the German Microcensus (Statistisches Bundesamt, 2005, 2007) the resulting samples were slightly biased with regard to the same socio-demographic char-

[1] This method of controlling for acquiescence assumes the existence of balanced pairs of true and false items per dimension.