Journal of Research in Personality 43 (2009) 335–344

Contents lists available at ScienceDirect

Journal of Research in Personality

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

Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally

Martin Bäckström *, Fredrik Björklund, Magnus R. Larsson

Department of Psychology, Lund University, Box 213, 22100 Lund, Sweden

ARTICLE INFO

Article history:
Available online 31 December 2008

Keywords:
Personality
Social desirability
Factor analysis
Big Five

ABSTRACT

The factors in self-report inventories measuring the five-factor model (FFM) correlate with one another although they theoretically should not. Study 1 showed, across three different FFM-questionnaires, that almost all of the common variance between factors can be attributed to a single general factor related to social desirability. In Study 2, simple rephrasing of items from a FFM-questionnaire made them substantially less socially desirable, while the inventory's empirical (five factor) structure remained the same. Participants low in social desirability showed little difference between how they responded to the original items vs. the neutral items. For participants high in social desirability the difference was considerably larger. The simplicity of reducing social desirability in self-rating inventories of the FFM, and the usefulness of this endeavor, is discussed.

1. Introduction

One of the basic assumptions of the five-factor model (FFM) is that major personality differences can be captured by five separate or independent dimensions (Hofstee, 2003). But many inventories designed to measure the model have revealed rather high correlations between factors (Block, 1995). The present research concerns this problem and investigates two issues that may elucidate why the correlations exist. The first issue is the possible existence of one or more higher-order factors causing the correlations. The second issue is the content of the uppermost factor(s) in the hierarchy, and if it can be attributed to the relative social desirability of the items used in inventories.

Adherents of the FFM have suggested that the five factors at the domain level are orthogonal (e.g. McCrae, Zonderman, Costa, Bond, & Paunonen, 1996, p. 552), i.e. that they are not correlated with each other. Why are there then often correlations between the five factors in inventories measuring the FFM? There have basically been two ways of answering this question:

1. The correlations are caused by nuisance factors related to how we measure personality (e.g. self-reports).
2. The correlations portray the factual relations between the traits.

Supporters of the first type of answer do not regard correlations between traits as a problem for the model, while adherents of the second type of answer question the validity of the five-factor model and have suggested modifications to it. One way to explain the correlations is that there exists a hierarchical structure, where factors at lower levels are connected with each other through their relationships with higher-order factors. Many models have been presented that attribute a substantial part of the correlation between personality factors to a general evaluative factor. This general factor has been recognized by Peabody and Goldberg (1989) and by Hendriks, Hofstee, and de Raad (2002) among many others, but it has an even longer history within personality measurement (e.g. Edwards, 1957). Models with more than one higher-order factor have also been proposed. Early in the history of the FFM some theoreticians suggested that one or more factors were really sub-factors to other factors at a higher level. For example, Eysenck (1992) proposed that Conscientiousness and Agreeableness are sub-factors of Psychoticism. Digman (1997) proposed two factors, alpha and beta (rotational variants of Eysenck's model) as higher-order factors to the Big Five. Most recently DeYoung, Peterson, and Higgins (2002) proposed, based on Digman's model, that facets of the five factors reveal a loading structure that can be attributed to two higher-order factors, called plasticity and stability. In an even more elaborated hierarchical model (Ashton, Lee, & Goldberg, 2004) it was suggested that there is a hierarchy traversing from one factor, over two factors, and so on ending up in five, six or seven factors.

If a general factor is a nuisance factor, what does it consist of? In a study by Edwards and Edwards (1991) it was shown that social desirability, the tendency to rate in the direction of the cultural norm, often appears in the first principal component when analyzing personality scales. This result has been replicated by Hendriks et al. (2002), Musek (2007), and Bäckström (2007), among others. Social desirability is one of the main concerns for personality

* Corresponding author.
E-mail address: martin.backstrom@psychology.lu.se (M. Bäckström).

0092-6566/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jrp.2008.12.013

measurement, and there is a huge amount of published studies on the subject (over 2000 articles in PsycINFO use the descriptor). Whereas most researchers agree that a social desirability factor exists, there is less agreement concerning its relation to personality. A rather heated debate has taken place between those who suggest social desirability to be a nuisance factor in personality testing and those who suggest it to be part of personality itself (Smith & Ellingson, 2002, for a review).

One way to tackle the problem of nuisance is to accept that personality inventories are influenced by social desirability and arrange to take care of this after the data have been gathered. For example, one may measure social desirability separately and then control for this factor statistically. Another way to control for social desirability is to use an ipsative response format, where the rater has to choose between items that are comparable in desirability. Both these methods have been criticized (e.g. Paulhus & Vazire, 2007). Some test developers have been reluctant to remove variance related to social desirability from the inventories since the criterion related validity for the inventory have been shown to decrease (McCrae & Costa, 1983).

Many have attempted to create inventories that are resistant to social desirability. For example, Jackson (1984; see also Jackson, Ashton, & Tomes, 1996) tried to reduce social desirability by careful selection of items. But such inventories have had a relatively weak impact on the field compared to the mainstream FFM inventories (e.g. NEO PI-R; Costa & McCrae, 1992), both as regards basic personality research and its applications. Inspired by the work of Jackson and by Peabody's (1967, 1987) work on the evaluative aspect of personality descriptors, we set out to investigate if the general factor can be minimized by rewriting the items in a way that does not activate as much social desirability, but where the factor structure is preserved.

To summarize, the general hypothesis is that correlation between personality factors at the domain level (the Big Five) can be largely attributed to a general factor, and that this general factor is caused by social desirability concerns activated by the semantic content of the test items.

2. Study 1

The first study deals with the number of higher-order factors in inventories that measure the FFM. Basically there are two different methods to reveal hierarchical structure in personality inventories; the Bottom–up method and the top–down method. The Bottom–up method takes items, facets or domains (e.g. the five factors) as a starting point, and tries to find higher level factors among them. If there are correlations between these factors, new factors at a higher level are extracted until only a single common factor is left or the factors do not correlate. This method is by far the most common one in personality research. The top–down method, however, starts with a single factor at the highest level, based on the correlation between items, facets or domains. If there is systematic covariance left after defining the first factor level, more factor levels are created based on the left-over covariance at the lower level. These factors are allowed to correlate with the factors at a higher level. In the present study the top–down method will be used, since the research question deals with the relative importance of different levels, particularly the most general factor level compared with the level below it in the hierarchy.

The prerequisite for hierarchical factor analysis is that there are correlations between factors at a lower level. In the present study, the lowest level that will be analyzed is the five domains that constitute the factors in the FFM. The specific question addressed is how many higher-order levels, if any, that can be found above the domain level using the top–down method. In other words, the present study differs from studies dealing with the factor structure of personality markers or personality tests in that we are not investigating the number of factors necessary to represent personality differences. Instead we have deliberately chosen to start from the FFM and inventories that measure this model. It is the correlation found between the factors at the domain level (the five factors) that is our main interest. Furthermore, we do not deal with the importance of the factors of the FFM. Instead we investigate the importance of the higher level factors of FFM inventories.

2.1. Method

2.1.1. Participants

Three datasets were used in the first study. All were gathered from the same internet site (http://www.pimahb.com). Most of the participants in the three datasets were unique participants, but there was approximately 20% participant overlap between them[1]. The first dataset consisted of 747 participants, 293 men and 454 women, tested with the IPIP-100 inventory. The second consisted of 1200 participants, 392 men, 748 women, and 60 who did not reveal their sex, tested with the IPIP–AB5C. The third consisted of 878 participants, 261 men and 547 women, and 70 who did not reveal their sex, tested with Goldberg's (1992) personality markers. In all three datasets the mean age was approximately 30 years (SD = 13). The age distribution was skewed to the left, with a dominance of participants in the lower age groups. All participants were recruited directly from the internet; they were spontaneous visitors to the web site.

2.1.2. Materials

Three five-factor inventories (one based on personality markers) were used in the study. All three inventories had the same five-point Likert scale response format, ranging from "not at all accurate" to "very accurate". The items were presented to the subject in a randomly generated order.

The IPIP-100 inventory used here is based on a Swedish translation of the short IPIP five-factor inventory (Bäckström, 2007; Goldberg, 1999; Goldberg et al., 2006). It has 100 items distributed among the five factors; 20 items per scale. The personality dimensions measured in the inventory are Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness. The reliability for the present materials, as indicated by Cronbach alphas, were .92, .88, .92, .93, and .86 for the Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness scales, respectively. These figures were comparable to those reported by Ashton and Lee (2005) for the English version[2].

The IPIP–AB5C (Goldberg et al., 2006) is a five-factor inventory developed by Goldberg and his collaborators at the IPIP, on the basis of the circumplex model suggested by Hofstee, de Raad, and Goldberg (1992). It consists of 486 items that are approximately equally divided on 45 sub-traits, nine for each of the five super-factors. The circumplex model states that many sub-factors of personality, e.g. personality-descriptive adjectives, have loadings on at least two super-factors. The inventory was translated by the first author and re-translated into English by a professional translator. Only minor differences were identified in this process. Bäckström

[1] Internet visitors sometimes take more than one inventory but most take only one. We chose not to exclude subjects taking more than one since we thought this would make the sample less representative.

[2] The Swedish version was compared with the English regarding factorial invariance, using two large samples (N = 5565 for the Swedish and N = 461 for the English version). The factor structure was found to be very similar (e.g. PNFI (Parsimony normed fit index) was 0.99 with group restriction and 0.87 without group restriction. In other words, a model defining the factor coefficients to be the same for the English and Swedish version was better than a model where the factor coefficients were allowed to differ).