Personality and Individual Differences 96 (2016) 31–35

Contents lists available at ScienceDirect

Personality and Individual Differences

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

# Is the general factor of personality based on evaluative responding? Experimental manipulation of item-popularity in personality inventories

Martin Bäckström *, Fredrik Björklund

*Department of Psychology, Lund University, Box 213, SE-221 00 Lund, Sweden*

ARTICLE INFO

*Article history:*
Received 21 November 2015
Received in revised form 12 February 2016
Accepted 22 February 2016
Available online xxxx

*Keywords:*
Self-ratings
Personality
Big Five
General factor of personality

ABSTRACT

The general factor of personality (GFP) is understood as a hierarchically superordinate factor, which suggests that it and the subordinate personality traits are mutually dependent on one another. If a personality inventory captures the subordinate traits the GFP should appear too. Likewise, manipulating the GFP should affect the subordinate traits and vice versa. The current study was an attempt to uniquely affect the size of the GFP by manipulating the evaluativeness of the inventory. First we estimated a general factor in a standard (evaluative) personality inventory, and found it to be robust. Then we estimated it in an inventory with evaluatively neutralized items, and found it to be unreliable. Finally, the neutralized inventory was made evaluative again. As expected, the GFP reappeared, suggesting the increased evaluative content to be the cause. Results are discussed in relation to personality assessment and to higher order factors in personality theory. It is suggested that for determining whether the GFP exists or not researchers should turn to other measures than personality inventories.


## 1. Introduction

The general factor of personality (GFP) is conceptualized as a higher-order factor causing some of the variation in lower-order personality traits (Musek, 2007; Rushton, Bons, & Hur, 2008; Van der Linden, te Nijenhuis, & Bakker, 2010). The correlation that exists between the different factors in personality inventories is seen as a consequence of the GFP by some (e.g. Musek, 2007; Van der Linden et al., 2010) but not by others (e.g. Revelle & Wilt, 2013; Riemann & Kandler, 2010).

Research on the GFP has focused on extracting GFP from personality inventories. The current study concerns precisely this; the role of the test instrument in GFP research. Notably, personality instruments vary in the extent that they give support to a GFP. In instruments such as Jackson's PRF (Jackson, 1984) and the HEXACO (de Vries, 2011) there is little general correlation between the main scales, whereas in other instruments there is substantial correlation. We propose that this is due to some instruments being under greater influence than others from individual differences in how respondents approach the inventory. Some people respond identically to items that refer to the same content, regardless of wording. Other people are more sensitive to item wording. Arguably, a major factor behind the degree of correlation between scales is the instrument's level of evaluativeness (Pettersson, Turkheimer, Horn, & Menatti, 2012), which can be defined as the extent to which the inventory affords responses that reflect the cultural norm for desired behavior. As indicated by previous research, the variation related to a GFP should diminish in the inventory if the items are made less evaluative (Bäckström, Björklund, & Larsson, 2009). However, to the extent that evaluativeness and support for a GFP are linked together, it should also be possible to redesign an inventory in the other direction. It should be possible to alter an instrument that shows little evidence of a GFP so that it, after the evaluativeness treatment, now does show evidence of a GFP. Making the inventory more evaluative should increase the correlation between the factors, which could effectively be seen as substantiating the general factor. Thus, the current study concerns how systematically increasing vs. decreasing the degree of evaluativeness in personality inventories, (while keeping other factors constant) affects the support for the GFP in the inventories.

How should we understand the nature of the general factor in personality inventories? The advocates of the GFP suggest that the factor constitutes a hierarchically superior content factor affecting all underlying factors in the model in the same way. In a five-factor model context, this implies that the GFP influences each of the five underlying content factors. According to this way of reasoning, if the inventory actually captures the lower-order personality traits, then it will capture the GFP too.

People who are high in GFP should tend to be higher on all factors, i.e. high in extraversion, agreeableness, conscientiousness, emotional stability and openness to experience. The standard five-factor model, where each factor is assumed to be independent of the other factors, is more parsimonious in the sense that there is one factor less, which also avoids the problem of where to place it in the factor structure. However, as has been shown several times, this is not a model that is generally supported by personality inventory data (Bäckström, 2007; Musek, 2007).

Following the reasoning above, if GFP is verified by correlation among the big five factors of inventories alone, all instruments that do

* Corresponding author.
*E-mail address:* Martin.Backstrom@psy.lu.se (M. Bäckström).

http://dx.doi.org/10.1016/j.paid.2016.02.058
0191-8869/

not show such correlation would weaken the case for the GFP. In other words, an inventory which captures the big five but has no higher order factor, should be problematic from a GFP perspective. If it is possible to affect the pattern of correlation between factors by applying a simple strategy, such as the evaluative neutralization method (where item popularity is manipulated; Bäckström & Björklund, 2013), exclusive reliance on self-reported personality data in investigating the GFP hypothesis may be problematic. It should be noted that the reduction of the GFP by means of e.g. the neutralization method could be caused by other factors than evaluativeness, e.g. reduction in reliability or independent changes of all the factors by introducing content from other personality content factors (which would need to be separate from the big five, otherwise they would create correlation according to the GFP theory).

To succeed with the removal of the GFP, the manipulation needs to affect items of all the sub-factors equally, otherwise the success would only be partial, e.g. there would be no general reduction in correlation or the factorial structure would break down. Granting these obstacles, this appears to be possible; in previous research using evaluative neutralization there has been no reduction of reliability and the factor structure has been intact. All five factors have been retained, while their intercorrelations decreased (Bäckström, Björklund, & Larsson, 2014). However, recreating the GFP by means of a reversal of the same method as was used to remove it would be an even more powerful demonstration of the influence of evaluativeness on GFP. This issue is to the core of the debate on the validity of the GFP as a personality content factor, and the main focus of the current study. In a similar vein as in a classic ABA-design study, we expect to be able to extract a robust GFP in the original (evaluatively loaded) version of a personality inventory, expect little evidence of a GFP after the inventory has been evaluatively neutralized, and a robust GFP again after it has been made evaluative again. We hypothesize that relatively small changes to the wordings of items affect their popularity and the correlation between scales that are based on the items. More precisely, we predict that reducing evaluativeness will mean that the GFP becomes smaller and increasing evaluativeness that it becomes larger. We also predict that the general factor that is expected to appear in the "re-evaluized" inventory will correlate substantially with the general factor of a typical Five factor model (FFM) inventory, which would be key with regard to concluding that it captures GFP-related variance. In other words, our hypothesis is that simple rephrasing of item to be more popular will influence the ratings of every item, from all the five scales, in the same way.

## 2. Method

### 2.1. Materials

This study is based on three different inventories measuring the FFM model.

#### 2.1.1. IPIP-NEO

The first one is the IPIP-NEO inventory from the International Item Pool (Goldberg et al., 2006). We used a 200-item Swedish version of this inventory which has been described elsewhere (e.g. Bäckström et al., 2014) and shown to be a valid instrument to measure the FFM, on par with the more well-known NEO-PI-R (Costa & McCrae, 1992).

#### 2.1.2. FFM-Neutralized

This 160-item inventory was developed in a project on evaluative neutralization of items from the IPIP-NEO. Validities are in the range of the original inventory (Bäckström et al., 2014). The factors of the inventory have lower intercorrelations than the IPIP-NEO (and the NEO-PI-R) and the facets have fewer cross-loadings than IPIP-NEO.

#### 2.1.3. FFM-Evaluative

This inventory was created for the sole purpose of the present study. It was based on the same items as the FFM-Neutralized and has four facets per factor. The items of the FFM-Evaluative were constructed iteratively by rephrasing all items from the FFM-Neutralized. They were rephrased to become more popular (as described in Bäckström and Björklund (2013)), i.e. made more evaluative in the sense that many participants in the population should find them attractive (and willing to rate high). For example, *Want to constantly meet and enjoy the company of friends and colleagues* was turned into *Want to have people around me*. The item *Carry out all tasks, even when I see them as unimportant* was turned into *Important tasks can sometimes be put on hold*, and *Have to achieve everything I set my mind to do* was turned into *Often achieve what I set my mind to do*.

An item was categorized as popular when the mean rating was .3 steps above the midpoint of the five point Likert scale (negatively worded items were reversed). To check whether the changes were successful, we had a fairly large sample of respondents (between 86 and 190) make self-ratings on the revised items. The rephrasing was iterated until most items were clearly popular (mean above 3.3).

The IPIP-NEO had 151 (out of 200) items that were popular, the FFM-neutralized had 20 (out of 160), and the FFM-Evaluative had 100 (out of 160). The item mean was 3.63, 3.03 and 3.41 for the IPIP-NEO, FFM-Neutralized, and FFM-Evaluative, respectively. In other words, although the FFM-Evaluative had items that were more popular, they were not as popular as the original IPIP-NEO.

### 2.2. Participants and procedure

Participants were Swedish-speaking spontaneous visitors to the site www.pimahb.se, i.e. they were not actively recruited for the study. All participants volunteered and were provided with some feedback on their results. Across samples there was about 65% women and the mean age was ca 30 years. Visitors who register on the site report their educational level and if they work. Of those who have registered, about 40% have reported more than three years of college level education, 21% college studies for less than three years, 31% have reported high-school and the rest a lower level of education (e.g. secondary school). Of all registered about 61% have reported working more than 20 h a week.

The IPIP-NEO was administrated separately, and items were presented in a random order. The FFM-Evaluative and FFM-Neutralized were administered at the same time, and items were presented together randomly.

#### 2.2.1. Statistical analyses

The hypothesis that evaluativeness brings about a general factor in personality inventories was tested with Confirmatory Factor Analysis. To extract the presumed general factor, we created a bi-factorial model where the FFM factors were defined by their five respective facets and the general factor, with all loadings fixed to 1, was defined as a common factor loading on all the 20 facets (see Fig. 1, panel 2). To test the hypothesis that making inventories more vs. less evaluative increases vs. decreases the GFP, we estimated two models for each inventory. The models included the Big Five factors that the inventories were designed to measure as well as a measurement factor with loadings to all subscales of the inventory. For each inventory, the first model defined the measurement factor to have zero correlation for all subscales (observed variables, equivalent to Fig. 1, panel 1), and the second model defined the measurement factor to have unit loadings for all subscales.

Since we are interested in the exact proportion of variance that can be attributed to the general factor in different inventories we used the Normed Fit Index, which measures the proportion of covariance explained by the models. We supplemented NFI with the Comparative Fit Index that adds a penalty for larger models and is more common in the literature. In addition, as an alternative way of estimating the amount of systematic variance in the general factor, the