



## Variance determines self-observer agreement on the Big Five personality traits

Jüri Allik<sup>a,\*</sup>, Anu Realo<sup>a</sup>, René Mõttus<sup>a</sup>, Tõnu Esko<sup>b</sup>, Janne Pullat<sup>b</sup>, Andres Metspalu<sup>b</sup>

<sup>a</sup>Department of Psychology, Estonian Center of Behavioral and Health Sciences, University of Tartu, Estonia

<sup>b</sup>Institute of Molecular and Cell Biology, Estonian Biocenter, University of Tartu, Estonia

### ARTICLE INFO

#### Article history:

Available online 3 June 2010

#### Keywords:

Personality ratings  
NEO PI-R  
Self-observer agreement  
Visibility of traits  
Assumed similarity  
Restriction for range

### ABSTRACT

It is widely believed that, on those personality traits that are more visible to an external observer, two judges will reach a higher level of agreement than on those traits that are more difficult to judge. This view is challenged in the current paper, using a sample of 672 participants in the age range of 18–87 years who described their own personality and were judged by an external observer who knew them well, using the NEO PI-3 questionnaire (McCrae, Costa, & Martin, 2005). The self-observer agreement on the 30 personality subscales varied from .38 (O3: Feelings) to .57 (E5: Excitement Seeking). Approximately one-half of the variance in the agreement level was explained by the standard deviation of the sum scores of these subscales: self-observer agreement was higher in the subscales on which individual differences were larger. After correction for the range of variance, differences in self-observer agreement substantially diminished. It is proposed that judges who know each other well reach an approximately equal level of agreement on all the Big Five personality traits.

Crown Copyright © 2010 Published by Elsevier Inc. All rights reserved.

### 1. Introduction

In order to function efficiently in the social environment, people need to regularly make personality judgments about themselves and other people. Typically, the accuracy of such judgments is estimated by the degree of agreement between ratings made by different judges. Although self-observer and between-observer agreement do not prove that personality ratings are accurate, it is a necessary prerequisite for their accuracy. When self-ratings are compared to those of well-informed observers, judges tend to achieve at least moderate cross-observer agreement for most personality traits (Funder & Colvin, 1997; Kenny, 1994). For example, the mean or median interobserver trait agreement among well-acquainted informants is almost invariably .40 or higher for all the Big Five personality dimensions (Connolly, Kavanagh, & Viswesvaran, 2007; Hall, Andrzejewski, Murphy, Schmid Mast, & Feinstein, 2008; McCrae et al., 2004). This level of agreement is not trivial considering the complicated chain of events required for an accurate personality judgment: the target of judgment must display behaviors and cues that are relevant to the trait being judged and the judge must detect these cues and correctly use them to make judgments (Funder, 1999; Funder & Colvin, 1997).

Considering the intricacies of personality judgments, it is not surprising that researchers are inclined to believe that this agreement between judgments can be easily compromised and that

there are several moderators that can substantially reduce self-observer agreement. One of these moderators which has received much attention is “judgability”. For instance, psychologically better adjusted individuals are easier to judge than less well adjusted people (Colvin, 1993; Furr, Dougherty, Marsh, & Mathias, 2007). Another equally powerful moderator is the ability to make correct personality judgments from available information: some individuals are believed to be better judges of personality than are others (Letzring, Wells, & Funder, 2006; Realo et al., 2003; Taft, 1955). However, “judgability” and “good judges” appear to be less influential than previously thought: as shown by Allik and colleagues (2010), self-observer agreement does not generalize easily from one personality trait to another since targets and raters from the same target–rater pairs may occupy identical or nearly identical positions in their respective rankings on one personality trait but can have a considerable disparity in their rankings on another personality trait.

In this paper, we focus on the question of whether some personality traits are easier to judge than the others and what might be the reason for this.

#### 1.1. Are some traits easier to judge than others?

Over the years, several different studies have shown that it is easier to reach self-observer agreement on some personality traits than on others. For example, numerous studies have shown that traits defining Extraversion are easier to judge than traits defining Neuroticism (Connolly et al., 2007; Funder & Dobroth, 1987; Hall et al., 2008; Park & Judd, 1989). Another recent meta-review

\* Corresponding author. Address: Department of Psychology, University of Tartu, Tiigi 78, Tartu 50410, Estonia.

E-mail address: [juri.allik@ut.ee](mailto:juri.allik@ut.ee) (J. Allik).

showed that the mean correlation in observer-ratings, corrected for coefficient alpha in self-ratings and interrater reliability, was .62 for Extraversion, which was higher than for all other traits (.51 for Neuroticism, .59 for Openness, .46 for Agreeableness, and .56 for Conscientiousness, respectively) (Connolly et al., 2007). In several other meta-reviews, Extraversion has invariably been shown to achieve higher self-observer agreement than Neuroticism (Hall et al., 2008). Despite such consistent findings, the between-trait differences in agreement are still relatively small. In fact, median or mean values of cross-observer correlations for all personality traits are basically within the same range. The median value of Extraversion (.47) for cross-observer agreement in studies using different measures of the Five-Factor Model with single raters was only slightly higher than for Neuroticism (.43), Openness (.43), Agreeableness (.40), and Conscientiousness (.41) (McCrae et al., 2004).

Researchers have proposed several explanations for why it is easier to reach agreement on some personality traits than on others.

#### 1.1.1. Visibility of traits

The concept of *visibility* has been the most widely used as an explanation for why some personality traits are easier to agree upon than others. This term encompasses several related concepts such as *judgability*, *confirmability*, *observability*, and *availability* of traits (Funder, 1995; Funder & Drobth, 1987; Gangestad, Simpson, DiGeronimo, & Biek, 1992; Paunonen, 1989; Tausch, Kenworthy, & Hewstone, 2007). For example, Funder and Drobth (1987) asked participants to rate 100 California Q-sort items on nine subjective dimensions. One of these dimensions was how easy it is to imagine specific, observable behaviors that would confirm or disconfirm a trait and another one was how easy it is to judge the degree to which another person had the trait. Six of these dimensions grouped into one factor, which was interpreted as reflecting each trait's "easy visibility" to an outside observer. Most visible traits belonged to the Extraversion domain, while the least visible traits were from the Neuroticism domain. The correlation between agreement scores and visibility ratings was moderately positive:  $r = .42, p < .001$  (Funder & Drobth, 1987). Later studies confirmed that better agreement is reached on more visible traits (Funder & Colvin, 1988). Thus, variation in self-observer agreement may be attributed to systematic differences in items' content.

#### 1.1.2. Assumed similarity

Assumed similarity refers to the tendency to perceive others as similar to oneself (Beer & Watson, 2008) or some generalized other (Cronbach, 1955). It seems logical to suppose, when trait information is not readily available, that judgments are made on the assumption that others are similar to oneself or some hypothetical, idealized person. The usual way to calculate assumed similarity is by correlating an individual's self-rating with the average of his or her ratings of each of the other group members (i.e., grouping within judges). As expected, Beer and Watson (2008) found that assumed similarity was statistically significant for Neuroticism ( $r = .32$ ) and near zero for Extraversion ( $r = -.07$ ). This seems to confirm the principle that more visible traits are judged on the basis of veridical information, while less visible traits are described according to a self-based heuristic (Beer & Watson, 2008). The degree to which raters' own personalities contributed to target ratings also correlated negatively with self-observer agreement ( $r = -.60$ ) (Beer & Watson, 2008; Ready, Clark, Watson, & Westhouse, 2000) and exhibited a strong relationship with the visibility of traits ( $r = -.73$ ). Thus, when an observer is asked to rate a target on difficult-to-judge traits, he or she is more inclined to project his or her own personality on the target.

Instead of relying on one's own personality, it is also possible to imagine someone who has socially desirable traits. It has been noticed that self-observer agreement is stronger on neutral rather than socially favorable personality traits (John & Robins, 1993). On socially more desirable personality traits, judges seem to be more guided by their expectations rather than actual information about personality traits which, in turn, could lead to lower self-observer agreement.

Unfortunately, it is impossible to calculate assumed similarity when a person has assessed only one target in addition to her- or himself. In many cases, it is neither economic nor possible to obtain ratings of multiple targets from one rater. Another problem is that well-acquainted samples have actually shown higher assumed similarity correlations than stranger samples (Beer & Watson, 2008; Kenny, 1994). This casts doubt on the universality of the assumption that the absence of appropriate information is always compensated for with assumed information about oneself or a hypothetical person. Nevertheless, assumed similarity seems to be, so far, the strongest moderator of self-observer agreement.

#### 1.1.3. Affectivity

Item visibility and assumed similarity are not the only known moderators of self-observer agreement. Watson and his colleagues (2000) noted that Neuroticism scales lead to considerably higher self-observer agreement than the PANAS negative affectivity scales ( $r = .46$  versus  $.29$ ), in spite of the fact that they both measure approximately the same content (Watson et al., 2000). The only obvious difference seems to be the format of items. While items in personality questionnaires, such as the NEO PI-R (Costa & McCrae, 1992), for instance, are formulated in short statement form (e.g., "feel inferior to others" or "have a low opinion of myself"), in the PANAS Negative Affect Scale, people are asked to rate how they have felt during a certain period of elapsed time or at the present moment, using single words such as *distressed*, *hostile*, or *angry*. Thus, it could be wording, format, or instruction, not visibility of traits alone, which determines these differences in self-observer agreement.

#### 1.2. Restriction of range

However, there may be purely statistical reasons why the correlation between two variables,  $x$  and  $y$ , can vary in its magnitude. A common problem facing researchers is calculating correlation in some population of interest on the basis of a restricted sample. It is well known that a sample correlation can deviate from a population correlation for a variety of reasons, including sampling error, measurement error, and restriction of range (Sackett & Yang, 2000). Usually the sampling problem is understood in terms of individuals: instead of the entire sample, a small fraction of individuals is available for investigation. Another domain in which the restriction of range may have some relevance is the selection of items for personality questionnaires. In the study of inclination towards artistic experiences, for example, researchers can devise such items as "I'm not really interested in the arts" (reversed item) or "Certain kinds of music have an endless fascination for me" (McCrae et al., 2005). For the study of altruistic dispositions, in turn, the following items have been successfully devised: "Some people think I'm selfish and egoistical" (reversed item) or "Most people I know like me" (Costa & McCrae, 1992). Although impeccable specimens, these items are but a small fraction from a large pool of items which can potentially be used to measure Openness to Aesthetics or altruistic dispositions. For this very reason, it is possible that a given sample of items is only measuring a restricted range of Openness or Altruism, compared to what could conceivably be covered by the whole imaginable set of items. Any restriction in the range possibly resulting in depressed variance in traits

Download English Version:

<https://daneshyari.com/en/article/951414>

Download Persian Version:

<https://daneshyari.com/article/951414>

[Daneshyari.com](https://daneshyari.com)