Valérie-Anne Ryser

Psychometric properties of extra-short Big Five personality measures in multi-topic surveys: Documenting personality traits in the SHP and MOSAiCH

FORS Working Paper series

The FORS Working Paper series presents findings related to survey research, focusing on methodological aspects of survey research or substantive research. Manuscripts submitted are papers that represent work-in-progress. This series is intended to provide an early and relatively fast means of publication prior to further development of the work. A revised version might be requested from the author directly.

Further information on the FORS Working Paper Series can be found on
www.forscenter.ch

How to cite this document:
Ryser, V.-A. (2015). Psychometric properties of extra-short Big Five personality measures in multi-topic surveys: Documenting personality traits in the SHP and MOSAiCH. *FORS Working Paper Series, paper 2015-5*. Lausanne: FORS.

Summary

Using three different Swiss nationwide datasets, one from MOSAiCH (2007) and two from the Swiss Household Panel (SHP-2009 and SHP-2014), the aim of this study is to assess the psychometric properties of two extra-short multi-item scales designed to assess the Big Five personality traits: the 10-item Big Five Inventory-Ten (BFI-10) and the 15-item Big Five Inventory-Short Version (BFI-15).

The measurement models of both scales were evaluated using confirmatory factor analyses (CFA). General result patterns demonstrated that the basic models based on the three datasets showed marginal fit. The two instruments do not adequately capture the Big Five personality traits. Neither the MOSAiCH sample nor the two SHP samples support the five-factor structure postulated by the theory of both scales.

However, the two instruments benefit largely from modifications: acceptable models can be found based on residual correlations and adaptation of the theoretical models, although they do not always measure all of the Big Five personality traits but only some of them.

Keywords: Personality traits, Big Five inventory, CFA, general population surveys, within country comparisons.

Psychometric properties of extra-short Big Five personality measures in multi-topic surveys: Documenting personality traits in the SHP and MOSAiCH

Valérie-Anne Ryser[1]

1. Introduction

Many general population surveys cover a wide range of topics and present general perspectives on social change, particularly the dynamics of changing living conditions. To understand how individuals react to changes such as life transitions or life events, self-reported measures of individual characteristics are needed. In this perspective, the assessment of different psychological aspects of human adaptation and personality traits is becoming a key component of such surveys. For these multi-topic surveys, interview time is limited and the length of the questionnaires is central; thus, short measures with acceptable psychometric properties are crucial. For this purpose, based on the Big Five personality traits theory and questionnaires, several so-called mini-markers of personality traits (Saucier, 1994) have recently been developed (e.g. Gerlitz & Schupp, 2005; Gosling, Rentfrow, & Swann, 2003; Rammstedt & John, 2007). The Big Five personality traits theory provides information about the differences between individuals on five principal personality dimensions, which are "Openness to Experience", "Conscientiousness", "Extraversion", "Agreeableness" and "Neuroticism" (see Table 1 for a description of each trait).

Using confirmatory factor analyses (CFA)[2], the aim of the present study is to investigate the psychometric properties of two different extra-short scales of the Big

---

[1] FORS – Lausanne
[2] Confirmatory factor analysis (CFA) is a kind of factor analysis that allows testing the hypothesis that a relationship exists between observed variables and their underlying latent constructs. CFA models deal with measurement models and are by nature hypothesis-driven (Brown, 2006).

Five personality traits: the 10-item Big Five Inventory-Ten (BFI-10; Rammstedt & John, 2007)[3] and the 15-item Big Five Inventory-Short Version (BFI-15; Gerlitz & Schupp, 2005)[4]. Our analyses will be based on three different randomly selected samples of individuals living in Switzerland: one from the MOSAiCH survey (Measurement and Observation of Social Attitudes in Switzerland) and two from the Swiss Household Panel (SHP). Each sample is composed of individuals living in the three main linguistic areas (the Swiss-German, French and Italian[5] parts) of Switzerland.

The goal is to test the psychometric properties and robustness of the construct of these two scales on a general representative population. The first part of the study involves the psychometric evaluation of the BFI-10 on two different surveys using different modes of data collection. We will first assess the psychometric properties of the BFI-10 on the MOSAiCH sample and then on the SHP survey. Given that the BFI-10 has already been validated in German (Rammstedt & John, 2007), and since the SHP sample is sufficiently large, it is possible to replicate the validation of the BFI-10 on the SHP-Swiss German sample and then conduct multi-group analyses to test measurement invariance across these different linguistic areas to test the validity of the scale in French and Italian. The goal is to then validate the extent to which the same construct is measured across three different linguistic groups because traits are considered universal and, therefore, whether the same construct is effective across contexts (McCrae & Terracciano, 2005). The second part of the paper involves evaluation of the psychometric properties of the BFI-15 on a subsample of the SHP sample.

The article is organized as follows: first, we will discuss the theoretical and empirical background of the Big Five personality trait questionnaires and the two different extra-short measures of the Big Five personality traits; second, we will present the MOSAiCH and SHP samples on which our study is based; and, third, using CFA and multi-group confirmatory factor analyses (MGCFA), we will demonstrate the validity and quality of the two Big Five short scales. The final section draws conclusions and discusses further implications and the limits of such short personality measures in broad topic surveys.

---

[3] See Annex 1 Table 1 for a presentation of the items.
[4] See Annex 1 Table 2 for a presentation of the items.
[5] There are actually four official languages in Switzerland—Swiss-German, French, Italian and Rhaeto-Romansch—but the SHP and MOSAiCH surveys do not conduct interviews in Rhaeto-Romansch. People who speak Rhaeto-Romansch also speak another language such as Italian or Swiss-German, which are languages of the SHP questionnaires.

2. The Big Five personality construct

The Big Five personality construct is a well-established frame of reference in the field of personality psychology (McCrae & Costa, 2003). It assesses how the respondent positions him- or herself relative to a list of statements and aims for a better understanding of the differences between individuals on five principal personality dimensions: "Openness to Experience", "Conscientiousness", "Extraversion", "Agreeableness" and "Neuroticism" (Table 1). Each of the five dimensions, defined in Table 1, is represented by six facets (McCrae & Costa, 2003).

However, the insufficient theorizing for some dimensions and the lack of universal consensus on the conceptual definition of each trait (Block, 2010) have led to an absence of agreement on the factor content. Some authors included items related to warmth in the "Extraversion" domain (The NEO questionnaire; Costa & McCrae, 1985), whereas others introduced them in the "Agreeableness" dimension (John & Srivastava, 1999). A trait is commonly defined as temporally stable and heritable, at least in part. Moreover, traits are considered universal (McCrae & Costa, 1997) and therefore effective across contexts (McCrae & Terracciano, 2005). Several authors have demonstrated a relative stability of personality traits during adulthood, even if they suggested that not all personality traits are equally stable (Srivastava, Gosling, & Potter, 2003).

Table 1: Description of the Big Five domains and their facets

Big Five dimensions	Definition	Facets and adjectives
Openness to experience	The tendency to be open to new aesthetic, cultural, or intellectual experiences.	Ideas (curious), fantasy (imaginative), aesthetics (artistic), actions (wide interests), feelings (excitable),values (unconventional)
Conscientiousness	The tendency to be organized, responsible, and hardworking.	Competence (efficient), order (organized), dutifulness (careful), achievement striving (thorough), self-discipline (not lazy), deliberation (not impulsive)
Extraversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability.	Gregariousness (sociable), assertiveness (forceful), activity (energetic), excitement-seeking (adventurous), positive emotions (enthusiastic), warmth (outgoing)
Agreeableness	The tendency to act in a cooperative, unselfish manner.	Trust (forgiving), straightforwardness (not demanding), altruism (warm), compliance (not stubborn), modesty (not show-off), tender-mindedness (sympathetic)
Neuroticism	Emotional stability is "predictability and consistency in emotional reactions, with absence of rapid mood changes." Neuroticism is "a chronic level of emotional instability and proneness to psychological distress."	Anxiety (tense), angry hostility (irritable), depression (not contented), self-consciousness (shy), impulsiveness (moody), vulnerability (not self-confident)

Notes: Adapted from John and Srivastava (1999).

The Big Five personality traits have been assessed through a large number of different instruments of varied lengths. Initially, each of the six facets of the five personality traits is measured by a set of questions concerning thoughts, feelings and behaviors. Consequently, a large number of items are necessary to provide an exhaustive measure of the Big Five personality dimensions. One of the most commonly used instruments to measure the Big Five dimension is the NEO PI-R (Personality Inventory–Revised), which is a paper and pencil questionnaire (Costa & McCrae, 1985). The NEO PI-R is a questionnaire composed of 240 to 249 items, depending on the version. However, the main limitations of this questionnaire are its length and the considerable amount of time and attention/concentration needed for its completion, which also imposes quite a high cognitive burden. In addition, such a questionnaire is

not well suited to interviewer-administered surveys such as CATI (computer-assisted telephone interviewing) as well as CAPI (computer-assisted personal interview) surveys. To allow the measurement of personality traits in multi-topic surveys with strong time constraints and different modes of data collection, several shorter versions of the Big Five have been developed.

The Big Five Inventory (BFI-44) (John, Donahue, & Kentle, 1991) consists of 44 items. However, even though it is much shorter than the NEO PI-R, it is still too long for multi-topic surveys. For this reason, some much shorter versions of the Big Five personality traits questionnaire have been recently provided. Derived from the BFI-44, two mini-markers have been developed: the Big Five Inventory-Short Version (BFI-15) (Gerlitz & Schupp, 2005), consisting of 15 items, and the Big Five Inventory-Ten (BFI-10), consisting of 10 items (Rammstedt & John, 2007). These scales are meant for research with strong time constraints. In the same vein, Gosling et al. (2003) developed the Ten-Item Personality Inventory (TIPI) composed of 10 adjectives to assess the Big Five personality traits.

Because all five factors cover a broad range of facets measured by a large number of items, shorter versions of the Big Five lead to the focus on a reduced number of components in order to decrease the number of items. According to Losoncz (2009), it means that these shorter instruments include items that are close to the prototypical cores of each typical factor. In addition, not all facets of the factor are represented. The benefit of this procedure leads to a more homogeneous scale despite a loss of accuracy of the measure of each factor.

3. Big Five mini-markers: Assessing the Big Five personality traits in multi-topic surveys

Currently, the measures of personality traits have not only aroused interest in psychology but also in economics (Almlund, Duckworth, Heckman, & Kautz, 2011) and political science (Gerber et al., 2011). To meet this demand, several multi-topic nationwide surveys have introduced mini-markers of the Big Five personality traits, for instance: the German Socio-Economic Panel (GSOEP); the British Household Panel Survey and the UK Household longitudinal Study (UKHLS) (https://www.iser.essex.ac.uk/bhps); the Household, Income and Labor Dynamics in

Australia survey (HILDA) (Lucas & Donnellan, 2009); the 2005 round of the International Social Survey Program (ISSP); MOSAiCH; and the SHP. Although several personality traits mini-markers have been introduced in a large number of population-based surveys, little is known about the psychometric properties of these different short- or extra-short scales for the general population.

John and Srivastava (1999) evaluated three different short instruments measuring the Big Five personality traits: the TDA (100-item trait descriptive adjectives) (Goldberg, 1992), which according to John and Srivastava (1999) is the most commonly used measure based on single items; the 60-item NEO-FFI - a shorter measure of the NEO PI-R by Costa and McCrae (1992); and the 44-item BFI (John, et al., 1991). Using confirmatory factor analysis (CFA), their evaluation is based on a sample of students. The results demonstrate that "Extraversion," "Conscientiousness," and "Neuroticism" are the most reliable measures, whereas "Agreeableness" and "Openness," tend to be less reliable. Conducting CFA on the most basic model (five uncorrelated latent factors and observed measures loaded onto these dimensions) leads to marginal fit. With the addition of method factors, the correlations of residuals allowing cross loadings lead to acceptable models.

Based on a sample of students, the TIPI (Gosling, et al., 2003) was evaluated. The results demonstrate that despite being a very short measure of personality traits compared to standard multi-item instruments, the TIPI reached adequate levels of convergent validity, test-retest reliability, patterns of predicted external correlates and convergence between self and observer ratings (Gosling, et al., 2003). The authors pointed out, however, that the TIPI is a "reasonable [proxy] for longer Big-Five instruments especially when research conditions dictate that a very short measure be used (Gosling, et al., 2003, p. 523)".

The BFI-10 has been validated in German and English on samples of students and dog owners (Rammstedt & John, 2007). This extra-short instrument shows retest reliability, structural validity, convergent validity with the NEO-PI-R and its facets and sufficient external validity using peer ratings. Using exploratory factor analyses, the Big Five structure has been replicated in this abbreviated set of items. In addition, norms and values have been provided based on a representative German population (Rammstedt, 2007).

The HILDA survey introduced two instruments to measure personality traits: the TIPI (Gosling, et al., 2003) and the Trait Descriptive Adjectives-40 (TDA-40) (Saucier, 1994). Based on the HILDA which is a general population survey, an analysis done by Losoncz (2009), using principal component analysis, demonstrated that two components have been extracted from the TIPI instead of the expected five, whereas the results from the TDA-40 extracted eight components. Thirty-three items fit quite well with the five-factor concept, but seven items formed an additional three components. Losoncz (2009) concluded that instruments used in the HILDA performed well enough to provide an adequate measure of personality traits.

Based on the GSOEP datasets, Lang and colleagues (2011) used exploratory structural equation modelling (ESEM) to evaluate the psychometric properties of the BFI-15. Lang and colleagues (2011) emphasized the robustness of the measure for self-rated personality dimensions in the population of young and middle-aged adults. For the elderly, the authors claimed that the mode of interview is particularly important. For this specific population, telephone interviews are associated with a significant cognitive load that negatively impacts the reliability of self-assessment of personality in this population.

Based on the SHP preliminary exploratory analyses (Reiser, 2011), CFA demonstrates that the SHP data do not support the Big Five factor model. These analyses questioned the response bias and the formulation of the questions. Descriptive and exploratory analyses of the BFI-10 (Reiser, 2011) raised several questions, and the theoretical structure of the BFI-10 was questionable.

To conclude, the results of these different studies, which assess the psychometric properties of different Big Five mini-markers, present some limitations. First, most of these Big Five short scales have been validated on samples of students, which means that these samples are homogeneous, with people that have the same age and level of education (e.g. Benet-Martínez & John, 1998; Gosling, et al., 2003; John & Srivastava, 1999; McCrae & Terracciano, 2005; Rammstedt & John, 2007). This limits the generalization of the findings, insofar as these samples are highly selective. Little is known about the validation of such scales based on the general population with a heterogeneous sample that does not share the same age, level of education or step in the life course. In addition, the results that assess the psychometric properties of these short or extra-short Big Five scales in the general population present mixed results about the psychometric properties of these instruments. Second, the wide use

of exploratory factor analyses (e.g., McCrae & Costa, 1997) raises some questions; based on such analyses, the five-factor structure of personality emerged. Yet, to our knowledge, exploratory factor analyses are not the most appropriate method to evaluate the psychometric properties of these scales. Exploratory factor analyses do not have an a priori hypothesis about the structure of the factors or patterns of measured variables, whereas the Big Five personality traits theory has a strong assumption about the structure of the scales. Finally, the general population-based studies that introduced a Big Five personality trait scale used somewhat different instruments composed of different sets of variables. This has led to some difficulties of comparability between these scales across surveys.

4. The present research

The main aim of the present study is to validate and test the psychometric properties of two personality trait mini-markers – the BF-10 (Rammstedt & John, 2007) and the BFI-15 (Gerlitz & Schupp, 2005) – on the general representative population using CFA. Testing the psychometric properties of these scales on representative sample of the general population, while these scales have most of the time been evaluated on samples of students or young adults, is the main strength of our study. The second strength is to test the measurement invariance of the model across language groups. The main purpose is to assess to what extent the measurement properties are equivalent in subgroups of the population and to determine if the same underlying construct is comparable across groups. Finally, we will compare these two mini-markers.

The study follows three main purposes: first, using CFA to provide a validation for general population samples living in Switzerland for the BFI-10 and BFI-15, based respectively on the MOSAiCH and SHP surveys, which are two datasets that were collected using two different collection modes. The second purpose is to conduct multi-group confirmatory factor analyses (MGCFA) based on three linguistic subsamples of the SHP to assess to what extent the BFI-10 measurement is equivalent across the three linguistic areas in Switzerland. Because the German version of the BFI-10 has already been tested, we conducted the first analyses with the sample composed of individuals speaking Swiss-German. Then we conducted multi-group confirmatory factor analyses to test the equivalence of the construct

between the three linguistic groups. This analysis is the prerequisite for the generalization of an instrument. Finally, the third aim is to compare the two different short personality measures – the BFI-10 and the BFI-15 – among the two samples of the SHP to determine the most parsimonious model.

5. Data, sample and method

5.1. Data and sample description

We based our investigations on three different datasets drawn from two Swiss nationwide surveys: one from MOSAiCH and two datasets derived from the SHP[6]. The participants of the three datasets are drawn from a probabilistic sample representing the population living in the three main linguistic areas in Switzerland (Swiss-German, Italian and French).

The MOSAiCH is a biennial cross-sectional survey that aims at a better understanding of the Swiss population's values and attitudes toward a wide range of social issues. Participants are aged 18 and above. The MOSAiCH is a computer-assisted and personal face-to-face interview (CAPI). The questionnaire covers a wide range of topics in the social sciences. The MOSAiCH survey has been carried out since 2005, but selected parts have been conducted in Switzerland for more than 20 years, including modules for the International Social Survey Program as well as selected questions from the Eurobarometer in the Switzerland survey. The Big Five was introduced in 2007 in the Swiss sample. Each statement of the BFI-10 must be evaluated on a scale ranging from 1 (*agree strongly*) to 5 (*disagree strongly*). For the purpose of our study, we used a subsample of MOSAiCH-2007, and a total of 772 individuals aged 18 to 65[7] (336 males / 436 females) were taken into account. Table 2 displays the descriptions of the sample.

The SHP is a computer-assisted telephone interview (CATI) survey. This longitudinal survey follows three random samples of households: A first sample has been drawn in 1999 (5,074 households with a total of 12,931 individuals). Due to attrition, a second

---

[6] The Swiss Household Panel (SHP) is based at the Swiss Centre of Expertise in the Social Sciences – FORS – located at the University of Lausanne. The project, devoted to analyzing the changing living conditions in Switzerland, is funded by the Swiss National Science Foundation (SNF).
[7] We selected individuals under 66 because we know that, for the oldest population, answering the BFI-10 by telephone can be difficult and could undermine the psychometric properties of the scale (Lang et al., 2011).

sample (2,538 households with a total of 6,569 individuals) and a third sample (4,093 households and 9,945 individuals) were drawn in 2004 and 2013, respectively. All household members older than 14 years of age are personally interviewed on a yearly basis. The SHP covers a broad range of topics and approaches in the social sciences. For the present study, we first used the sample from 1999 and the refreshment sample from 2004. The BFI-10 was introduced in the SHP questionnaire in 2009. For the sake of comparability across different surveys, the translated versions of the BFI-10 scale in the SHP are the same as for the MOSAiCH survey. The Swiss-German version of the BFI-10 has been adapted from the translation done by Rammstedt and John (2007) for the CATI interview. Initially, each statement of the BFI-10 was to be evaluated on a scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Since the standard response scales of the SHP questionnaires are from 0 to 10, we adapted the 5-point Likert scale and labels of the BFI-10 to an 11-point scale ranging from 0 (*completely disagree*) to 10 (*completely agree*).

The analysis was conducted on a total of 5,360 individuals aged 18 to 65[8]: 1,369 living in the French-speaking, 3,796 in the German-speaking and 195 in the Italian-speaking part of Switzerland. Table 2 presents the description of the socio-demographic characteristics of the sample and shows that women are somewhat overrepresented. According to Lipps (2007), "although attrition is comparatively high in the SHP, it is not particularly selective with respect to important socio-demographic or -economic variables (Lipps, (2007, p. 63)".

The BFI-15 (Gerlitz & Schupp, 2005) was introduced in a subsample of the SHP in 2014. We first introduced the BFI-15 on a subsample of the SHP to test the psychometric properties of this scale, which is one of the aims of the present study, before introducing this scale to the entire SHP sample. In total, 241 participants of the SHP in 2014 wave 16, aged 18 to 65 (Table 2), answered the BFI-15. The German version of the BFI-15 was taken and adapted from the GSOEP, and the French and Italian versions were translated from the German version. As for the BFI-10, we adapted the BFI-15 to an 11-point scale ranging from 0 (*completely disagree*) to 10 (*completely agree*). The second analyses are based on this SHP subsample from wave 16 in 2014 composed of 241 individuals.

---

[8] We selected individuals aged 18 because of the comparison with MOSAiCH, where the younger participants are aged 18 years old.

Table 2: Socio-demographic characteristics of the MOSAiCH and the SHP samples

Samples	SHP; 2009				SHP; 2014 [2]	MOSAiCH; 2007
Linguistic areas	**French**	**German**	**Italian**	**Total**		
N	1369	3796	195	5360	241	772 [3]
N Gender						
Male	600	1680	79	2359	110	336
Female	769	2116	116	3001	131	436
Age groups / percentages						
18-34	26.4	24.9	13.8	24.9	18.7	25.1
35-49	40.0	36.9	48.7	38.1	30.3	41.6
50-65	33.5	38.2	37.4	37.0	51	33.3
Level of Education / percentages						
Low level	14.9	14.2	23.1	14.7	55.2	11.1
Middle	48.1	51.2	48.7	50.3	6.6	51.5
High	36.7	34.5	27.7	34.08	4.6	37.4
Type of occupation / percentages						
Full time	42.8	43.9	39.	43.5	45.2	49.2
Part time	29.7	30.7	28.2	30.4	32.8	30.5
Unemployed	2.3	10.1	2.6	1.5	2.5	0.9
Training	9.6	10	5.1	9.7	3.7	3.3
At home	15.2	14	24.1	14.6	14.9	13.2
Marital status / percentages						
Married, registered partnership	59.2	56.6	67.2	57.7	59.7	54.3
Single	27.2	31.5	20	30	28.2	27.3
Widowed	1.8	1.8	2.1	1.8	2.1	2.6
Divorced, separated [1]	11.8	10.2	10.7	10.7	9.9	14.6

Note: [1]) registered partnership dissolution; [2]) Because the Big Five items were only collected in the first phase of the fieldwork for the SHP-2014, there are many missing values on the education variable. In addition, respondents in the first phase of the fieldwork tend to have a lower level of education. [3]) 204 French; 533 Swiss-German; 34 Italian.

5.2. Analytical Strategy

We first conducted descriptive statistics to assess the mean and standard deviation for each item. The aim is to show the distribution of the population on each item. Second, we measured the internal consistency for each of the five personality trait dimensions. The aim is to control the items that measure the same trait producing similar scores. For this purpose, for the BFI-10, we tested the strength of the association (correlations) between the items, which are supposed to assess the same latent construct in order to evaluate to what extent the items that represent each trait fit well together. Since, for the BFI-10, each latent construct is assessed by two items, we then expect that the two components of each trait are moderately to strongly associated. For the BFI-15, three items composed each trait. For this reason, we computed Cronbach's alpha to measure the internal consistency of each trait. Cronbach's alpha value provides a measure of the reliability of each scale trait as well as an estimation of the proportion of the total variance, which is the common variance. If three indicators are supposed to measure the same latent variable, they should consequently share a certain amount of variance. Therefore, we expect moderate to strong internal consistency, as already presented in the literature (e.g., Benet-Martínez & John, 1998; John & Srivastava, 1999). Finally, we go a step further by conducting CFA to assess the factor structure of the observed variables. The aim of CFA, a particular form of factor analysis, is to verify to what extent the observed data fit the BFI-hypothesized measurement model. The general idea of CFA is to assess the relationship between observed measures or indicators such as questionnaire items and latent variables, also called factors or constructs, which are not directly measurable (Brown, 2006). We used CFA because the Big Five personality theory is very clear about the structure of the short scales developed. Based on past evidence and theory, five latent factors (Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness) are measured by a defined number of observed items. In addition, CFA includes the examination of the stability or invariance of the factor model over groups (Brown, 2006). The BFI-10 and the BFI-15 are both intended to be administered on a general population sample; therefore, we should establish that their measurement properties are equivalent in subgroups of the population and that the same underlying construct is comparable across groups. Because the SHP is composed of three linguistic groups, CFA makes it possible to test the invariance across these three groups.

The analyses were conducted using the software R and the following packages: Lavaan (Rosseel, 2012, 2015), semPlot (Epskamp, 2013) and semTool (Pornprasertmanit et al., 2014). Several indicators are considered to assess the model's quality of fit. Multiple indicators of fit are needed because they provide different information about model fit (for example, absolute fit, fit adjusting for model parsimony and fit relative to a null model; some are sample size-dependent). The use of a set of fit indices provides a more conservative and reliable evaluation of the model (Brown, 2006). According to Brown, "the goal of goodness-of-fit approaches is to identify the solution that reproduces the observed correlations considerably better than more parsimonious models (i.e., models involving fewer factors) but is able to reproduce these observed relationships equally or nearly as well as more complex solutions (i.e., models with more factors) (p. 30, 2006)."

Model fit is classically indicated by a non-significant *chi-square* test that shows the difference between observed and expected covariance matrices. Consequently, the smaller the values, the better the model fit. However, the chi-square test presents some limitations because it is dependent on the sample size: the chi-square increases with a larger sample size, which results in a higher probability of rejecting the model. For this reason, in addition to the chi-square, we used the *Root Mean Square Error of Approximation* (RMSEA), which is a fit index that avoids sample size issues. RMSEA assesses the extent to which a model fits a population reasonably well. Values very close to 0 suggest good model fit, and it has been argued that values close to .06 indicate a reasonably good model fit (Hu & Bentler, 1999). The classical cut-off criteria suggests that RMSEA values less than 0.08 can be considered as adequate; values less than 0.05 suggest good model fit; and RMSEA values higher than 0.1 should be rejected (Browne, Cudeck, Bollen, & Long, 1993). Finally, we also considered the *Comparative Fit Index* (CFI) and *Trucker-Lewis Index* (TLI). Both compare the target model with the null model. Larger values indicate better fit, which means that values greater than .95 are considered a good model fit (e.g., Brown 2006).

6. Results

6.1. Descriptives

Descriptive results (means, standard deviations and correlations/Cronbach's alpha) are shown in Table 3. Results from the descriptive statistics in Table 3 demonstrated

that, in the three samples (MOSAiCH, SHP-2009 and SHP-2014), the level of "Neuroticism" is quite low and individuals tend to be extraverted, open, agreeable and conscientious. Results are comparable with previous studies.

Looking at the strength of the associations between the pairs of items that make up each of the five personality trait dimensions, for the MOSAiCH sample, all the correlations[9] are significant; in addition, moderate to strong correlations are observed for the following dimensions: "Openness" (r = 0.21), "Conscientious" (r = 0.23), "Extraversion" (r = 0.41) and "Neuroticism" (r = 0.32). For the "Agreeableness" dimension, the correlation between the two items is weak (r = 0.12).

For the SHP-2009 in the German-speaking sample, all the correlations are significant, but the effect size between the two items of the "Agreeableness" dimension is negligible (r = 0.09; p < 0.001). For the French sample, the correlations between the two items from the "Agreeableness" dimension are negligible and non-significant (r = 0.02; ns). This means that there is no association between the two items or the association is negligible. For the Italian sample, many correlations are weak or negligible and also non-significant. This is the case for the following dimensions: "Openness" (r = 0.01; ns), "Extraversion" (r = 0.09; ns) and "Agreeableness" (r = 0.01; ns).

---

[9] Cut-off criteria for correlations: $r \geq$ | 0.40 | or higher means a strong relationship. r $\geq$ | 0.30 | to $r \leq$ | 0.39 | means a moderate relationship. r $\geq$ | 0.20 | to r $\leq$ | 0.29 | means a weak relationship. Finally, r $\geq$ | 0.0 | to r $\leq$ | 0.19 | means no or a negligible relationship.

Table 3: Descriptive statistics for the BFI-10 for the MOSAiCH and SHP samples

I see myself as someone who…	MOSAiCH sample 2007 (N=772)			PSM sample 2009 - Linguistic areas (N tot=5360) French (N=1'368)			German (N=3'795)			Italian (N=195)		
	M	SD	r	M	SD	r	M	SD	r	M	SD	r
Openness			.21***			.20***			0.24***			.01
has an active imagination	2.86	.84		6.79	1.82		6.98	2.01		7.15	1.77	
Has few artistic interests [2]	3.30	1.13		5.99	2.62		5.56	2.80		5.42	2.71	
Conscientiousness			.23***			.30***			0.26***			.186**
does a thorough job	3.31	.56		8.26	1.35		8.33	1.19		7.98	1.76	
tends to be lazy [1]	3.90	.92		6.55	2.40		6.74	2.29		5.86	2.53	
Extraversion			.41***			.39***			0.38***			.09
is reserved [1]	3.28	1.10		5.0	2.81		6.39	2.60		3.63	2.61	
is outgoing, sociable.	2.98	.75		6.84	1.89		7.63	1.59		7.52	1.78	
Agreeableness			.12***			.02			0.09***			.10
is generally trusting	2.91	.64		6.98	1.78		7.20	1.66		7.46	1.71	
tends to find fault with others [1]	3.71	.90		6.13	2.06		6.56	2.01		5.18	2.32	
Neuroticism			.32***			.30***			0.40***			.27***
is relaxed, handles stress well [1]	2.67	.89		6.74	1.84		7.10	1.67		6.65	2.10	
gets nervous easily.	3.40	1.04		5.71	2.42		6.23	2.24		5.18	2.58	

Notes: Item in the MOSAiCH are coded: 1 'Agree strongly' to 5 'Disagree strongly' whereas items in the SHP are coded 0 'Disagree strongly' to 10 'Agree strongly' [1]) items reversed in valence. [1]) In the SHP items which have been recoded; higher score means higher score on the latent construct. [2]) In the SHP this item is "has artistic interests". Mean (M), Standard Deviation (SD) and Pearson's correlation (r) ** $p \leq .01$; *** $p \leq .001$.

Table 4 demonstrates the results for the BFI-15. For this scale, each of the five traits – Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism – is measured by three items. For this reason, we computed Cronbach's alpha to measure the internal consistency of each trait. According to the literature (e.g. Benet-Martínez & John, 1998; John & Srivastava, 1999), we expect moderate to good internal consistency for each trait because the items measuring each dimension should be homogeneous. However, the results in Table 4 display some different outcomes: The Cronbach's alpha from the "Conscientiousness" dimension (α = 0.56) and the "Agreeableness" dimension (α = 0.42) is weak according to the usual criteria[10] and previous results (e.g. Benet-Martínez & John, 1998; John & Srivastava, 1999).

The conclusions of the descriptive results highlight some limitations. To sum up, in the MOSAiCH sample as well as in the SHP-2009 sample, the "Agreeableness" dimension shows a low level of internal consistency measured by weak correlations between the two items that composed it. In a context where the Big Five personality traits theory postulates high correlations, weak or negligible correlations and/or non-significant correlations suggest that the scale is not performing as expected. In the BFI-15 scale, the Cronbach's alpha weak scores demonstrate that for the "Conscientiousness" dimension as well as the "Agreeableness" traits, the internal consistency is low, which means that the amount of variance that is shared is lower compared to previous research (e.g. Benet-Martínez & John, 1998; John & Srivastava, 1999; Losoncz, 2009).

---

[10] The internal consistency is considered unacceptable if the Cronbach's alpha value is below .60, undesirable if between .60 and .65, minimally acceptable if between .65 and .70 and respectable between .70 and .80.

Table 4: Descriptive statistics of the BFI-15

PSM Subsample 2014	N= 240		
I see myself as someone who…	M	SD	*a*
Openness			0.68
has an active imagination	6.88	2.02	
has artistic interests	5.80	2.57	
is original	6.33	1.75	
Conscientiousness			0.56
does a thorough job	8.33	1.53	
tends to be lazy [1]	6.80	2.45	
is efficient	7.84	1.28	
Extraversion			0.72
is reserved [1]	4.48	2.26	
is outgoing, sociable.	7.02	1.99	
is communicative	7.64	1.84	
Agreeableness			0.42
is rude	6.20	2.49	
is forgiveness	7.25	1.98	
is kind	8.01	1.30	
Neuroticism			0.68
is relaxed, handles stress well	3.71	1.98	
gets nervous easily. [1]	4.64	2.44	
worries	4.73	2.27	

Notes: Items in the SHP are coded 0 'Disagree strongly' to 10 'Agree strongly'. Mean (M), Standard Deviation (SD) and Cronbach's alpha (α). [1]) Items reversed in valence.

6.2. Confirmatory factor analyses: replication of the theoretical model

Big Five Inventory-Ten (BFI-10): The aim of the CFA is to test the extent to which the observed data fit the BFI hypothesized measurement model. The first set of CFA is intended to replicate the theoretical basic model based on the data from the three datasets.

MOSAiCH: Based on the subsample of MOSAiCH, we move to the CFA to test the psychometric properties of the BFI-10 and estimate to what extent the observed items measured the latent theoretical Big Five construct. We replicate the theoretical model

based on the theoretical assumptions. As shown in Model 1, Table 5, the five latent constructs – Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism – are estimated through two items. First, a factor loading higher than 1.00 (imagination = 1.039) is observed. Second, low factor loadings (fault = 0.31; communicat. = 0.38) are identified. This means that the amount of variance in these two observed variables explained by the latent construct "Agreeableness" is low. For instance, a factor loading of 0.38 leads to 14% of the variance in the observed variable being explained by the underlying construct "Agreeableness". Third, the model shows marginal fit: the fit indices are mediocre to poor, which is not acceptable according to the usual criteria (Chi-square = 125.202, dl = 25, p > 0.001; CFI = 0.85; TLI = 0.73; RMSEA = 0.073).

Table 5. Theoretical CFA models based on the MOSAiCH and the SHP samples; Standardized coefficients

		MOSAiCH	PSM_2009		PSM_2014	
		Model 1	Model 2	Model 3	Model 4	
			German sample	French sample		
Latent variables	Items BFI-10[1]				Items BFI-15[1]	
Openess						
	imagination	1.039	1.2	0.889	imagination	0.847
	art. interest	-0.197[2]	0.193	0.226	art. interest	0.477
					original	0.674
Consienciousness						
	thorough job	0.424	0.522	0.768	thorough job	0.631
	lazy	-0.545	0.493	0.398	lazy	-0.383
					efficient	0.809
Extraversion						
	reserved	0.579	0.432	0.336	reserved	-0.497
	extravert	-0.699	0.876	1.158	talkative	0.775
					sociable	0.812
Agreeableness						
	communicat.	0.381	0.277	0.302	rude	0.287
	fault	-0.312	0.349	0.084	kind	-1.088
					forgiving	-0.19
Neuroticism						
	relaxed	0.486	0.747	0.609	worries	0.638
	nervous	-0.669	0.539	0.489	nervous	0.728
					remain calm	-0.582
Fit indices		Model 1	Model 2	Model 3	Model 4	
	Chi-Square/ddl	125.2/25	729.1/25	326.285/25		221.8/80
	CFI	0.853	0.828	0.797		0.826
	TLI	0.735	0.691	0.635		0.772
	RMSEA	0.073	0.086	0.094		0.086

Notes: [1]) items are summarized by one adjective. [2]) negative scores indicate items reversed in valence.

Finally, Table 6 demonstrates that the two latent constructs, "Agreeableness" and "Conscientiousness," present a very high correlation (r = 0.991). To conclude this first set of analyses, the MOSAiCH data do not support the five-factor structure of the BFI-10.

Table 6: MOSAiCH. Correlations of the five latent constructs

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Openness	1				
Conscientiousness	0.246	1			
Extraversion	-0.294	-0.465	1		
Agreeableness	0.186	0.991	-0.73	1	
Neuroticism	0.082	0.486	-0.332	0.945	1

SHP-2009: We move to the SHP-2009 sample to conduct CFA in order to test to what extent these data support the Big Five factor structure of the scale. Model 2, Table 5 displays the theoretical model for the BFI-10 scale. Results from Model 2 are based on the German sample, and results from Model 3 are based on the French sample.

As shown in Model 2, Table 5, results derived from the German sample demonstrate three main problems. The first problem is that the factor loading between the item "imagination" and the latent construct "Openness" is higher than 1.0, which is problematic. A second problem results from the low factor loading from the items composed of the "Agreeableness" dimension. Third, the fit indices are mediocre to poor (Chi-Square = 729.064, df 25, p < 0.001; CFI = .83, TLI = .69, RMSEA = .08) according to the usual criteria. Finally, a problem arises from the correlation of the two latent constructs "Conscientiousness" and "Agreeableness," which is higher than one (See Table 7).

Table 7: German sample (SHP-2009). Correlations of the five latent constructs

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Openness	1				
Conscientiousness	0.16	1			
Extraversion	0.30	0.44	1		
Agreeableness	0.12	1.17	0.56	1	
Neuroticism	0.19	0.53	0.35	0.69	1

CFA based on the French sample (Model 3, Table 5) leads to similar pattern of problems. Low factors loadings for the "Agreeableness" dimension; second Heywood case (i.e. factor loading higher than 1.0) on the "Extraversion" dimension and third, marginal fit indices. Finally, the high correlations between the two latent constructs "Agreeableness" and "Conscientiousness" (Table 8) and the dimensions "Agreeableness" and "Neuroticism" are largely problematic.

Table 8: French sample (SHP-2009). Correlations of the five latent constructs

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Openness	1				
Conscientiousness	0.265	1			
Extraversion	0.324	0.188	1		
Agreeableness	0.715	1.107	0.574	1	
Neuroticism	0.241	0.361	0.095	1.904	1

The last CFA based on the Italian sample did not converge. It is possible that the small sample size and lack of correlations for three dimensions play a role.

Based on the German and the French samples, we computed multi-group confirmatory factor analyses. However, the models show many problems and did not converge.

Big Five Inventory-Fifteen (BFI-15): Moving on to the psychometric properties of the BFI-15 based on the subsample of the SHP 2014, we conducted CFA to first test the theoretical model. For the BFI-15 scale, all five traits – Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism – are assessed through three items

each. Model 4, Table 5 displays the numbers for the theoretical model. Model 4, Table 5 demonstrates that we did not succeed in the replication of the theoretical model because different problems appeared. We note that the factor loading of the item "being kind" with the latent construct "Agreeableness" is greater than 1.0. Second, the factor loading of the item "forgiving" is particularly low (< 0.35). Finally, the different model fit indices used present mediocre scores (Chi-square = 221.816, df = 80; p < 0.001; CFI = 0.83; TLI = 0.77; RMSEA = 0.09).

6.3. Conclusion of the CFA analyses

To conclude this first set of CFA analyses that tested the construct of the BFI-10, neither the MOSAiCH sample nor the SHP samples supported the five-factor structure. Several limitations have been raised by the analyses. First, some indicators show a factor loading higher than 1.0, which is problematic. A standardized loading larger than 1 and a negative error variance indicate model specification error or problems with the sample or model-implied matrices (e.g., a non-positive definite matrix) or small sample size (Brown, 2006). Brown added, however, that "completely standardized factor loading above 1.0 may be admissible in such models, although this result might be indicative of multicollinearity in the sample data (Brown, 2006, p. 187)." A second problem is the low factor loadings (< 0.35) for some items: The square of a factor loading provides the amount of variance in the observed variable, which is explained by the latent factor. Consequently, low factor loading for the item "generally trusting" and "tend to find fault with others" (Table 5) means that the amount of variance in the observed variable explained by the latent construct "Agreeableness" is low. Table 5, Model 2 leads to 12% of the variance in the observed variable being explained by the underlying construct "Agreeableness." In addition, low factor loadings can lead to non-convergence or more likely to negative error variance (Gagne & Hancock, 2006). A third problem lies in the very high correlation and/or correlation higher than 1 between two latent constructs. This result indicates that, for the BFI-10, two latent constructs ("Agreeableness" and "Conscientiousness") are nearly indistinguishable. Based on previous results, admittedly conducted on the BFI-44 (John et al., 1991), this result is surprising. Finally, because the results of the BFI-10 are based on two different collection modes and lead to the same pattern of results, we presume that collection modes do not play a key role in the analyses.

6.4. Discussion of the CFA results

The main questions that arise are why the data do not support the theoretical model and the reasons why. Looking in the literature, several adjustments to the statistical model can be made. For instance, John and Srivastava (1999) used different methods to improve their model, such as adding a method factor; they also conducted modifications based on the residual matrix. Based on a theoretical perspective, the following aspects can be considered in further analyses.

The first aspect that we can consider is that, in both the BFI-10 and BFI-15, some of the items are negatively worded, whereas others are positively worded. These different wordings could undermine the psychometric properties of the scale. Following Marsh (1996), this issue can be controlled for by using the correlations of the residuals. We can control this aspect by correlating the residuals from the items that are negatively worded as well as correlating the residuals of the items that are positively worded.

Second, an acquiescence bias has been identified (Rammstedt & Kemper, 2011; Rammstedt, Kemper, & Borg, 2013) – which refers to the tendency to say "yes" to items independently of the item's content – that could undermine the psychometric properties of this scale. However, the addition of a methodological factor (Billiet & McClendon, 2000) could lead to a better model by controlling this acquiescence bias. In addition, a different kind of standardization could also be used to control this bias and then lead to an acceptable model. However, we can wonder whether this option is really relevant for the SHP, taking into account that the scale was numbered from 0-10 and only the end-points were labelled. In addition, other response tendencies (Fischer & Milfont, 2010) could also explain why the data do not support the Big Five structure. Response tendencies can be defined as the tendency for an individual to always choose the same scope of the scale. It can be an issue with the SHP because the BFI-10 introduced in the SHP-2009 is an eleven-point scale. According to Fischer and Milfont (2010), response tendencies can be controlled by using within-individual standardization.

6.5. Confirmatory factor analyses results: Adaptation of the theoretical model

The aim of the second set of CFA analyses is to control the extent to which the theoretical basic model, which showed marginal fit, would benefit from modifications

to improve its psychometric quality. Following John and Srivastava (1999) and based on the theoretical approaches presented above, the second set of analyses was then done to compute more complex models.

MOSAiCH. Based on the MOSAiCH sample, a model with modifications showed acceptable psychometric properties, as shown in Figure 1. We added a methodological factor, in line with Billiet and McClendon (2000). Beside the item "artistic interest," which shows a very low factor loading (–0.23), the other factor loadings are acceptable (> 0.35). The "Agreeableness" dimension was removed due to low factor loadings and high correlation with the "Conscientiousness" latent dimension. Finally, several correlations of the residuals had to be added in order to obtain an acceptable model. This new model presented good fit indices (Chi-square = 30.976, df = 12, p < 0.05; CFI = .96, TLI = .91, RMSEA = .046).

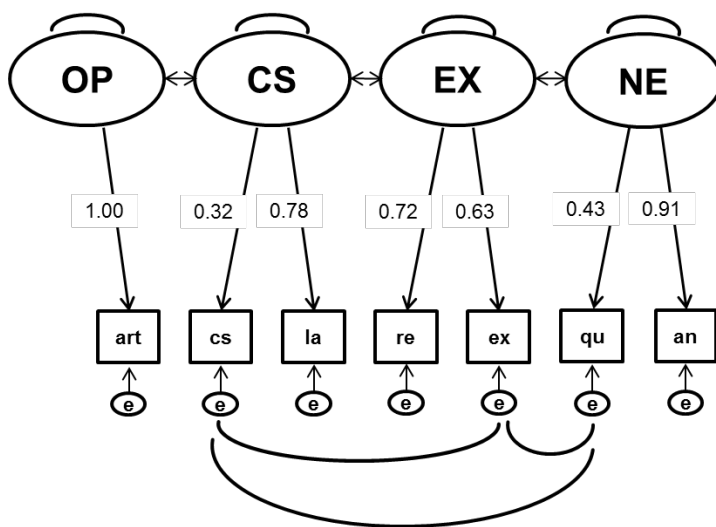**Figure 1. Four factor model. Adaptation of the theoretical model (based on MOSAiCH).**
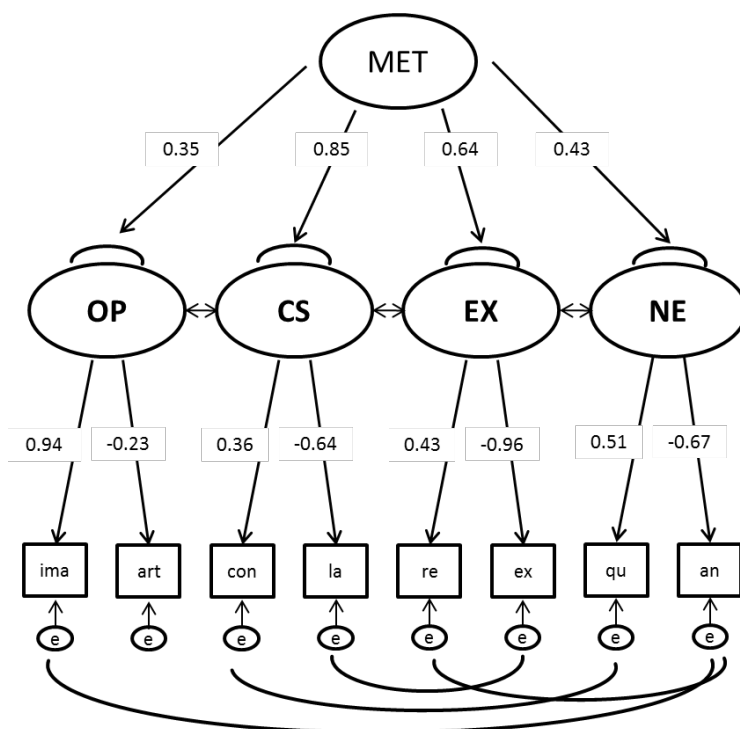
Notes: Standardized coefficients are reported.
Fit indices: Chi-square= 30.976, df= 12, p< 0.05; CFI=.96, TLI=.91, RMSEA=.046
Op= Openness; Cs= Conscientiousness; Ex= Extraversion; Ne= Neuroticism. The first letters mean the item linked with each trait "ima" refers to "active imagination"; "art" refers to "artistic interest"; "con" refers to "thorough job"; "la" refers to "tend to be lazy"; "re" refers to "reserved"; "ex" refers to "outgoing, sociable"; "qu" refers to "relaxed, handle stress well"; "an" refers to "gets nervous easily".

SHP-2009. Based on the German sample of the SHP data, we first removed the items with factor loadings higher than 1.0; second, following Marsh (1996), we correlated some residuals; and third, we removed the "Agreeableness" dimension, which shows low factor loadings compared to the one with the "Conscientiousness" dimension. This new model (Figure 2) presents good fit indices (Chi-square = 91.336, df = 43, p < 0.05; CFI = .96, TLI = .94, RMSEA = .043).

Figure 2. Four factor model. Adaptation of the theoretical model (based on the Swiss-German speaking sample, SHP 2009).
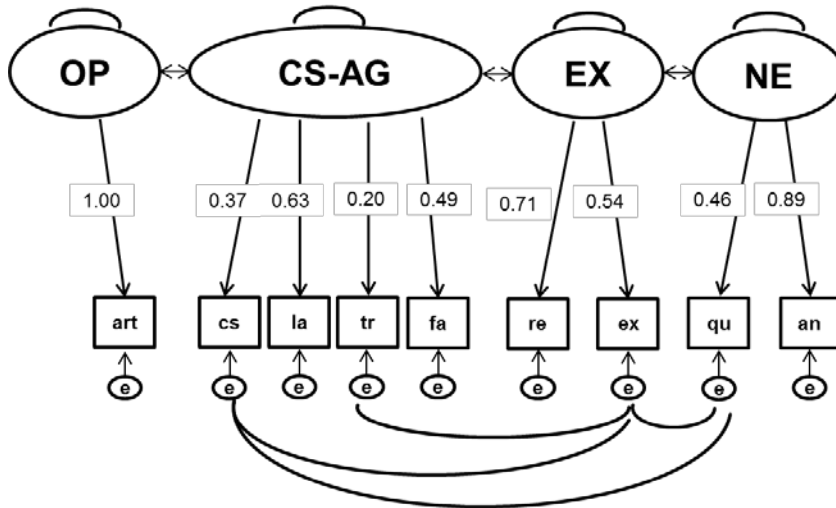

Notes: Standardized coefficients are reported.
Fit indices: Chi-square= 91.336, df= 43, p< 0.05; CFI=.96, TLI=.94, RMSEA=.043
Op= Openness; Cs= Conscientiousness; Ex= Extraversion; Ag= Agreeableness; Ne= Neuroticism. The first letters mean the item linked with each trait "art" refers to "artistic interest"; "cs" refers to "thorough job"; "la" refers to "tend to be lazy"; "re" refers to "reserved"; "ex" refers to "outgoing, sociable"; "qu" refers to "relaxed, handle stress well"; "an" refers to "gets nervous easily".).

Figure 3 displays the statistical measures for an alternative model. In this model, "imagination" was removed from the "Openness" latent construct, since it had a factor loading higher than 1.0. We computed one latent construct with "Conscientiousness" and "Agreeableness" – the two latent constructs with high correlations. Then, we correlated the residuals of some of the items. According to the cut-off criteria, the model fit is good (Chi-Square = 155.157, df = 48, p < 0.001; CFI = .95, TLI = .91, RMSEA = .046). However, from a theoretical perspective, it is not ideal to compute one latent construct composed of "Agreeableness" and "Conscientiousness." In addition, with this solution, there is a loss of information because two dimensions are mixed up. The five-factor structure is no longer observable; instead, we have a latent construct assessed by various items.

Figure 3 Four factor model. Adaptation of the theoretical model (based on the Swiss-German speaking sample, SHP 2009).

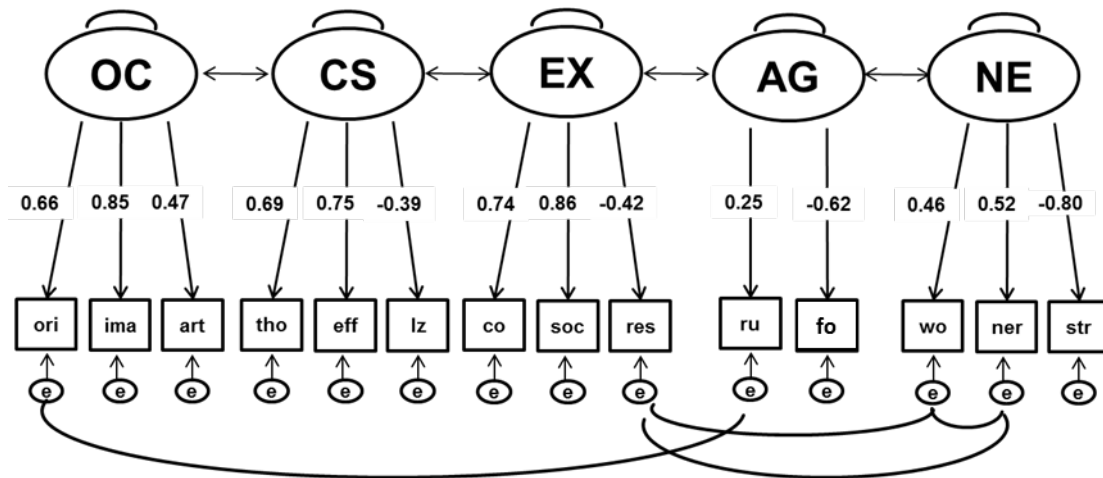Notes: Standardized coefficients are reported.
Fit indices: Chi-Square= 155.157, df= 48, p< 0.001; CFI=.95, TLI=.91, RMSEA=.046
Op= Openness; Cs= Conscientiousness; Ex= Extraversion; Ag= Agreeableness; Ne= Neuroticism. The first letters mean the item linked with each trait ("art" refers to "artistic interest"; "cs" refers to "thorough job"; "la" refers to "tend to be lazy"; "re" refers to "reserved"; "ex" refers to "outgoing, sociable"; "tr" refers to "generally trusting"; "fa" refers to "tend to find fault with others"; "qu" refers to "relaxed, handle stress well"; "an" refers to "gets nervous easily").

Based on the French sample, an alternative model could not be found without losing a large part of the theoretical structure of the Big Five. For this reason, in this working paper, we have not provided any solution.

SHP-2014/BFI-15. The second step was to understand the extent to which the theoretical model would benefit from modifications. After having removed the "sometimes rude to others" item, which had a factor loading higher than 1.0, and following Marsh (1996) with the correlations of some residuals, a second model (Figure 4) presented good fit indices (Chi-square = 91.336, df = 43, p < 0.05; CFI = .96, TLI = .94, RMSEA = .043). However, one item shows a very low factor loading with the latent construct "Agreeableness."

Figure 4. Adaptation of the theoretical model of the BFI-15 (based on the SHP subsample 2014)

Notes: Standardized coefficients are reported.
Fit indices: Chi-square= 91.336, df= 43, p< 0.05; CFI=.96, TLI=.94, RMSEA=.043
Op= Openness; Cs= Conscientiousness; Ex= Extraversion; Ag= Agreeableness; Ne= Neuroticism. The first letters mean the item linked with each trait ("ori" refers to "is original"; "ima" refers to "active imagination"; "art" refers to "artistic interest"; "tho" refers to "thorough job"; "eff" refers to "does things efficiently"; "lz" refers to "tends to be lazy"; "co" refers to "is talkative"; "soc" refers to "outgoing, sociable"; "res" refers to "reserved"; "ru" refers to "sometimes rude to others"; "fo" refers to "has a forgiving nature"; "wo" refers to "worries a lot"; "ner" refers to "gets nervous easily"; "str" remains calm in tense situation).

Conclusions of the second set of CFA. The disadvantage of these alternative models is that some information is lost. We removed items that showed factor loadings higher than 1.0; for the BFI-10, we removed the "Agreeableness" dimension because the items that were supposed to measure this latent construct performed badly (factor loadings < 0.38). For the BFI-15, the five-factor structure of the Big Five personality traits was kept since we correlated a certain number of residuals.

7. Conclusion

The aim of this study was to examine the psychometric properties of two extra-short measures of the Big Five personality traits – the 10-item Big Five Inventory-Ten (BFI-10; Rammstedt & John, 2007) and the 15-item Big Five Inventory-Short Version (BFI-15; Gerlitz & Schupp, 2005) – using three different Swiss nationwide datasets: one from MOSAiCH (2007) and two from the Swiss household panel (SHP-2009 and SHP-2014).

The first series of descriptive results highlight the lower level of internal consistency among these mini-markers, compare to instruments composed of a larger number of items. This result is in line with previous studies (e.g. Benet-Martínez & John, 1998; John & Srivastava, 1999). In addition, for the "Agreeableness" dimension, the associations between the items that measure this dimension were weak for both scales, in all of the samples. Items that were supposed to measure the same trait did not produce similar scores, which lead to weak internal validity for this dimension. The "Agreeableness" dimension tended to be less reliable compared to the other dimensions. Finally, for the BFI-15, in addition to the "Agreeableness" dimension, which showed low internal consistency, the "Conscientiousness" dimension also presented low internal consistency. These descriptive results mean that some items that were supposed to measure the same thing converged only slightly on the same construct.

Moving on to the CFA, several limitations can be shown in the second series of analyses, which tested the underlying Big Five personality construct. The general result patterns demonstrated that the basic models based on the three datasets showed marginal fit with the theoretical model. Neither the MOSAiCH sample nor the SHP samples supported the Big Five factor structure of the extra-short Big Five personality mini-markers considered in our study. CFA show several problems that tended to indicate problems with the factor structure of the BFI-10 as well as of the BFI-15. From a theoretical perspective, the problems indicated that none of the data support the five-factor personality structure. In addition, the results of the CFA reflected poor reliability for the "Agreeableness" dimension for both scales, meaning that in all of the datasets considered in the analyses, the items from the "Agreeableness" dimension did not seem to assess a consistent latent factor. The items did not seem to measure the core or prototypical aspects of the same underlying dimension and only poorly reflected the "Agreeableness" latent construct.
The last series of results emphasized the strong correlation of the two latent factors "Agreeableness" and "Conscientiousness" for the BFI-10, which indicated that these two latent factors were nearly indistinguishable. This is not in line with previous literature (e.g. John & Srivastava, 1999).

Several reasons could explain why the five-factor structure had encountered difficulties in the datasets considered by the analyses. First, such Big Five trait scales are often built and validated on samples of students that are homogeneous samples (with the same age and level of education); however, the aim of our study was to

evaluate the psychometric properties of these mini-markers in a general population survey, in which large heterogeneity is observed (for instance, in terms of age or level of education). Consequently, these mini-markers often seem to be sample dependent. A second reason that could explain why the five-factor structure was not identified by the datasets is the lack of theorization of certain personality dimensions. Block (2010) pointed out the atheoretical nature of the five factors and the lack of consensual understanding about personality traits and their content: the items related to warmth are included in either the "Extraversion" or the "Agreeableness" dimensions (John & Srivastava, 1999), depending on the authors.

However, the BFI-10 and the BFI-15 benefit greatly from modifications that improve their psychometric quality. Following John and Srivastava (1999), models can be improved using correlations of the residuals. The Heywood case can be solved after removing loadings higher than 1.0; the addition of the correlations of the residuals lead to an improvement of the model, although there is a loss of information or partial information. However, the use of different modes of centration and standardization have been used to try to diminish the different problems (Rammstedt & Kemper, 2011; Rammstedt, et al., 2013), but without success.

Finally, our last aim was to test the measurement invariance of the BFI-10 based on the SHP-2009 data. Unfortunately, we did not succeed in testing the measurement invariance between the three linguistic groups, even though acceptable models were found for the different linguistic groups.

Based on the MOSAiCH and SHP samples, we might conclude that two items per construct are not enough to identify the five-factor structure of personality traits in a general population. Hence, the strength of the BFI-15 is that, with three items per dimension, this scale comprises the minimum number of items per dimension for the identification of each personality trait. Considering the BFI-15, the model has to be adapted in order to have good model fit and fit indices that are reliable. We noticed a loss of information, but we found the general five-factor structure of the model, after some modifications.

Insofar as the assessment of the psychometric properties of the BFI-10 and the BFI-15 is based on samples derived from a general population, this study presents some limitations. Because of the nature of the surveys on which the analyses were based, it is impossible to assess the generalizability and consistency of these two scales over

time. Likewise, the external validity, as provided by peer ratings, is impossible to assess. Finally, convergent validity means that the correlations between the BFI-10 and the BFI-15 with other personality trait instruments are not possible to assess. These additional analyses, which are common when Big Five instruments are evaluated (e.g. John & Srivastava, 1999), could have given more accurate pictures of the adequacy of the BFI-10 and the BFI-15 in assessing the Big Five.

Finally, despite the marked popularity aroused by these scales in various domains, we can ask what they measure effectively. According to our results, they do not seem to always measure the five personality traits based on the Big Five theory but, instead, a certain number of traits. To avoid misinterpretation of the results, a study based on Big Five personality mini-markers should thoroughly examine the structure of the mini-markers before conducting further analyses. Researchers first have to disentangle whether the personality mini-markers measure five personality traits or a set of traits, which are occasionally unspecified.

8.   References

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). Personality psychology and economics, Discussion paper series // Forschungsinstitut zur Zukunft der Arbeit, No. 5500, http://nbn-resolving.de/urn:nbn:de:101:1-201104113733.

Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*(3), 729.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*(4), 608-628.

Block, J. (2010). The five-factor framing of personality and beyond: Some ruminations. *Psychological Inquiry, 21*, 2-25.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. 2006. *Confirmatory factor analysis for applied research. xiii*.

Browne, M. W., Cudeck, R., Bollen, K. A., & Long, J. S. (1993). Alternative ways of assessing model fit. *Sage Focus Editions, 154*, 136-136.

Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory: Manual, form S and form R*: Psychological Assessment Resources.

Epskamp, S. (2013). semPlot: Path diagrams and visual analysis of various SEM packages' output. *R package version 0.3, 2*.

Fischer, R., & Milfont, T. L. (2010). Standardization in psychological research. *International Journal of Psychological Research, 3*(1), 88-96.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research, 41*(1), 65-83.

Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., Raso, C., & Ha, S. E. (2011). Personality traits and participation in political processes. *The Journal of Politics, 73*(03), 692-706.

Gerlitz, J.-Y., & Schupp, J. (2005). Zur Erhebung der Big-Five-basierten persoenlichkeitsmerkmale im SOEP. *DIW Research Notes, 4*, 2005.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment, 4*(1), 26.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory: Versions 4a and 54, Institute of Personality and Social Research. *University of California, Berkeley, CA*.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). *Second Edition* (2nd edition). New York: Guilford.

Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the Big Five: robust across survey methods except telephone interviewing. *Behavior research methods*, 1-20.

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica, 47*(3), 111-121.

Lipps, O. (2007). Attrition in the Swiss Household Panel. *Methoden-Daten-Analysen, 1*(1), 45-68.

Losoncz, I. (2009). Personality traits in HILDA. *Australian Social Policy No. 8*, 169.

Lucas, R. E., & Donnellan, M. B. (2009). Age differences in personality: Evidence from a nationally representative Australian sample. *Developmental Psychology; Developmental Psychology, 45*(5), 1353.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810.

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509.

McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective*: Guilford Press.

McCrae, R. R., & Terracciano, A. (2005). Personality profiles of cultures: aggregate personality traits. *Journal of Personality and Social Psychology, 89*(3), 407.

Pornprasertmanit, S., Miller, P., Schoemann, A., Rosseel, Y., Quick, C., Garnier-Villarreal, M., et al. (2014). Package 'semTools'.

Rammstedt, B. (2007). The 10-item Big Five Inventory: Norm values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment, 23*(3), 193.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A ten-item short version of Big Five Inventory in English and German. *Journal of Research in Personality 41*, 203-212.

Rammstedt, B., & Kemper, C. (2011). Measurement equivalence of the Big Five: Shedding further light on potential causes of the educational bias. *Journal of Research in Personality, 45*(3), 121-125.

Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting Big Five Personality Measurements for Acquiescence: An 18-Country Cross-Cultural Study. *European Journal of Personality, 27*(1), 71-81.

Reiser, S. (2011). *Mehrdimensionale Analyse des Fünf-Faktoren Modells der Persönlichkeitseigenschaften*. Zürich: Soziologisches Institut.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Rosseel, Y. (2015). The lavaan tutorial.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of personality assessment, 63*(3), 506-516.

Srivastava, S., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*(5), 1041-1053.

Annex 1.

Table 1. Items of the Big Five Inventory-Ten adapted from Rammstedt and John (2007)

Personality trait Latent Variable	Question I see myself as someone who
Extraversion	… is reserved [1].
Agreeableness	... is generally trusting.
Conscientiousness	... does a thorough job.
Neuroticism	... is relaxed, handles stress well [1].
Openness	... has an active imagination.
Extraversion	... is outgoing, sociable.
Agreeableness	... tends to find fault with others [1].
Conscientiousness	... tends to be lazy [1].
Neuroticism	... gets nervous easily.
Openness	... has artistic interests.

Notes: [1]) Items reversed in valence.
Each trait is measured with two items.

Table 2. Items of the Big Five Inventory-Fifteen adapted from Lang et al. (2001)

Personality trait Latent Variable	Question I see myself as someone who
Extraversion	… is reserved [1].
Agreeableness	....is considerate and kind to almost everyone
Conscientiousness	... does a thorough job.
Neuroticism	... remains calm in tense situations [1].
Openness	... has an active imagination.
Extraversion	... is outgoing, sociable.
Agreeableness	... is sometimes rude to others [1].
Conscientiousness	... tends to be lazy [1].
Neuroticism	... gets nervous easily .
Openness	... values artistic, aesthetic experiences.
Extraversion	…is talkative
Agreeableness	… has a forgiving nature
Conscientiousness	… does thing efficiently
Neuroticism	… worries a lot
Openness	… is original, comes up with new idea

Notes: [1]) Items reversed in valence.
Each trait is measured with three items.