



An Investigation into the Factor Structure of a German Version of Goldberg's Big-Five Factor Markers

Nancy Borngraeber

Supervised by: Andy Bell

March 2010

An Investigation into the Factor Structure of a German Version of Goldberg's Big-Five Factor Markers

ABSTRACT

Previous studies have suggested the cross-cultural generalisability of a five-factor structure for personality traits. In this article, the utility of a translated version of the 50-item IPIP Big-Five factor markers was investigated in identifying five broad personality dimensions in a German context. A sample of German-speaking participants ($N = 236$) used the translated German version of the 50-item IPIP Big-five factor markers to describe themselves. Exploratory Factor Analysis demonstrated a five-factor structure that was nearly identical to the American structure. The reliabilities of the five factors were high except for the measure of Openness which was slightly lower but within acceptable limits. The homogeneity of the scales was analysed using confirmatory factor analysis to test the fit of a single-factor model. All scales were homogeneous with all of the items on each scale tapping a single factor. In line with previous research, the results of this study supported the generalisability of the five-factor IPIP structure in a German context. However, the analysis highlighted a number of items which needed to be discarded and should be revised to improve scale development of this measure.

KEY WORDS:	PERSONALITY	SCALE DEVELOPMENT	FIVE-FACTOR MODEL	IPIP	GERMAN
------------	-------------	-------------------	-------------------	------	--------

Introduction

Research in personality psychology has been a long-lasting discipline and it appears that a consensus is beginning to emerge in the form of the Five-factor model (FFM). In recent years, researchers and practitioners have increasingly employed the FFM based on evidence for the criterion-related validity of scores on FFM measures such as the NEO Personality Inventory-Revised (NEO-PI-R; Costa & McCrae, 1989) and the Big Five Inventory (BFI; John *et al.*, 1991). As such, studies have demonstrated the validity and utility of the FFM in predicting variables such as job performance (e.g. Barrick & Mount, 1991) and linking personality traits to physical health (e.g. Friedman *et al.*, 1995). However, “an emerging consensus is not the same as universal agreement” (Goldberg, 1995, p36) and more validating research will be necessary before critics of the FFM may be convinced.

As early as 1961, Tupes and Christal acknowledged that the difficulties in finding a consensus in the field of personality research was partly due to a lack of studies comparing the factor structures from similar sets of variables across diverse samples or procedures. Almost 40 years later, Goldberg (1999) has argued that “the science of personality assessment has progressed at a dismally slow pace” (Goldberg, 1999, p7). He attributed the seeming lack of progress to the fact that most of the broad-bandwidth personality inventories are proprietary instruments. As a consequence of the policies and practices of commercial inventory publishers, researchers require permission from the copyright holders and are charged for each questionnaire used, thereby hampering the use of comparative validity studies. Goldberg therefore proposed an international collaboration to develop an easily available, broad-bandwidth personality inventory which could be accessed by researchers at no cost. The International Personality Inventory Pool (IPIP) was thus created, allowing researchers to freely use the items in the pool and disseminate their findings to improve these.

Many popular psychological assessment instruments, originally developed in English, have been translated into numerous languages and are now commonly used throughout the world (Schmitt *et al.*, 2007). It is therefore not surprising that translations of the IPIP Big-Five factor markers – an increasingly popular personality measure found on the IPIP website - are available in more than 10 languages (Goldberg *et al.*, 2005). However, some researchers have expressed concern about the translations of psychological assessment instruments because of their underlying assumption that the core psychological constructs assessed by the measure are found across language and culture (e.g. Cheung & Leung, 1998). The analysis of the validity of translations seems therefore imperative, as comparing the assessment results from different cultures may otherwise become highly problematic (Brislin, 1993).

In relation to the FFM, previous studies have suggested the cross-cultural generalisability of a five-factor structure for personality traits, and large-scale studies of personality traits across cultures and languages suggested a robust factorial structure of personality measures such as the NEO-PI-R (McCrae, 2002) and the BFI (Schmitt *et al.*, 2007). Interestingly, however, there are only few reports on the characteristics of the IPIP Big-Five factor markers in other languages and cultures.

Mlačić and Goldberg (2007) investigated the 100-item and the 50-item versions of the IPIP Big-Five markers in both self-reports and peer ratings in Croatia. Similarly, Zheng *et al.* (2008) analysed the utility of the aforementioned versions of the IPIP Big-Five factor markers in both heterosexual and homosexual samples in China. Both studies showed clear five-factor orthogonal structures that were nearly identical to the American structure. The present study aims to extend this past work by investigating the factorial structure of the 50-item IPIP Big-Five factor markers in a German context. In what follows, some background on the FFM, the development of the IPIP Big-Five factor markers as well as a rationale for using an internet based study is provided to create a foundation for the present study.

Background Information

The Five-Factor Model of Personality

The development of the FFM is rooted in the lexical approach to personality. The lexical hypothesis posits that most of the socially relevant and salient personality characteristics have become encoded in the natural language (e.g. Allport, 1937). Allport and Odbert's (1936) identification and classification of trait words that describe individual differences provided some initial structure for the personality lexicon but offered little practical value in regards to offering a systematic framework for distinguishing, ordering and naming individual differences in people's behaviour and experience (John, 1989). Subsequent efforts to categorise these traits by Cattell (1947), Fiske (1949) and others resulted in the discovery and clarification of five distinctive dimensions. This five-factor structure has been replicated by a number of researchers (e.g. Norman, 1963) and eventually became known as the 'Big Five' (John & Srivastava, 1999). Although the title may be misleading, the Big Five structure does not imply that personality differences can be reduced to only five traits. Rather, these five dimensions represent personality at the broadest level of abstraction with each dimension subsuming a large number of more specific and distinct personality characteristics.

Researchers differ in their preference for factor labels, but the five dimensions of the FFM of personality can be termed as follows: Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to Experience (O). In brief, Extraversion involves attributes such as sociable, active and affectionate. Neuroticism contrasts emotional stability, as such individuals who score high on this factor are considered to be emotional, insecure and with a tendency to worry. Openness to Experience involves the extent to which individuals are curious, imaginative, unconventional, original and creative. Agreeableness refers to individuals who are trusting, helpful, forgiving and warm. Finally, Conscientiousness includes reliability, organisation, ambition and efficiency.

The Development of the Big-Five Factor Markers

The IPIP contains alternative versions of widely used personality inventories including the NEO-PI-R (Costa & McCrae, 1992), 16 Personality Factor Questionnaire (16PF: Conn & Rieke, 1994), California Psychological Inventory (CPI: Gough & Bradley, 1996), and the Hogan Personality Inventory (HPI: Hogan & Hogan, 1992). However, in addition to these public domain equivalents of proprietary

scales, the IPIP also contains a number of items known as the Big-Five factor markers (Goldberg, 1992).

Of interest in the present study is the 50-item IPIP measure of the Big-Five factor markers (Goldberg, 1992; see Appendix 1), although longer measures of the FFM comprised of IPIP items also exist. This scale was developed from Goldberg's (1992) 100 unipolar Big-Five factor markers which derived from early lexical studies in English. Research into these trait-descriptive adjectives suggested five broad factors and indicated high correlations between corresponding scales of the NEO-PI (Goldberg, 1992). However, Goldberg (1999) proposed that transforming these adjectives into short verbal phrases (e.g., 'Am the life of the party') would provide more contextual information than single words while still being more concise than items in many other inventories. A further benefit of the IPIP item format is that it should be easier to translate short behaviour-descriptive phrases than single trait-descriptive adjectives (Hendriks *et al.*, 1998). However, whilst there is some published work on the adjective markers and the IPIP versions of proprietary instruments, there appears to be little published research relating to the Big-Five factor markers (Gow *et al.*, 2005) and only a few reports on the characteristics of these markers in other languages (Mlačić & Goldberg, 2007; Zheng *et al.*, 2008).

Internet Research

The broad aim of this study is to investigate the factorial structure of a German version of a public domain personality inventory. To accomplish this goal a large and diverse sample of German speaking individuals was required and it was deemed that the most effective way to recruit suitable participants was through the internet. Data collection through the World Wide Web has increased in popularity over the past few years (Fraleay, 2007). The appeal of using the internet as a medium of data collection instead of the traditional pencil-and-paper methods can be easily understood. This method permits access to samples beyond the reach of methods typically used in psychological research (Ramírez-Esparaza *et al.*, 2006), allowing researchers to study people in a way that is relatively independent of location (Fraleay, 2007).

A common preconception about Internet research is that samples are not representative of a population and are instead predominantly young, affluent and male (Fraleay, 2007). Gosling *et al.* (2004) analysed large samples of people who completed measures of the Big Five traits over the internet and compared them to samples reported in the 2002 volume of the *Journal of Personality and Social Psychology* (JPSP). Their findings indicated that although internet samples were not representative of the population, they were generally more diverse as other samples typically used in psychological research which in many cases rely on university students (in 85 percent of the research reported in JPSP).

Although one should be cautious about the possibility of careless data namely participants submitting responses multiple times, Fraleay (2007) argued that similar caution should be taken in traditional methods when participants may be unmotivated to read the questions and instructions carefully and consequently provide inaccurate responses.

Method

Participants

The sample consisted of 236 German-speaking participants of which 67.4% ($N = 159$) were women and 32.6% ($N = 77$) were men. The age of the participants ranged between 18 to 84 years ($M = 33.08$, $S.D. = 12.17$) and the majority of 79.7% ($N = 188$) were German, followed by 14.8% ($N = 35$) who were Austrians and 5.5% ($N = 13$) were of other nationalities including Swiss, German-British, German-American, British, Polish, French, Italian, Russian, Montenegrin and Georgian. It should be noted that nationality was not a requirement of participation in this research as long as participants judged themselves as comprehending German to a good standard. Country of residence was similarly versatile and 44.9% ($N = 106$) were currently living in Germany, 28.4% ($N = 67$) in the United Kingdom, 15.7% ($N = 37$) in Austria, 8.5% ($N = 20$) in the United States of America and 0.4% ($N = 1$) in Switzerland, Poland, Netherlands, Argentina, Georgia and Montenegro respectively. The majority of participants attended university (53.4%, $N = 126$), 28% ($N = 66$) attended Gymnasium¹, 13.1% ($N = 31$) attended Realschule, and 5.5% ($N = 13$) attended Hauptschule. 58.1% ($N = 137$) stated being currently employed, 34.7% ($N = 82$) were students, 5.1% ($N = 12$) were pensioners, and 2.1% ($N = 5$) stated being currently unemployed. Table 1 gives an overview of the demographic characteristics of the sample.

¹ The German and U.K school systems are fundamentally different, therefore the following is provided only as a guide. There are three primary types of schools in Germany's secondary education: Gymnasium = grades 5-13; Realschule = grades 5-10; Hauptschule = grades 5-9

Table 1Demographic Characteristics of the Sample (*N* = 236)

Variables	Frequency	Valid %
Gender		
Female	77	32.6
Male	159	67.4
Nationality		
German	188	79.7
Austrian	35	14.8
Other (Swiss; German-British; German-American; British; Polish; French; Italian; Russian; Montenegrin; Georgian)	13	5.5
Country of Residence		
Germany	106	44.9
Austria	37	15.7
United Kingdom	67	28.4
United States of America	20	8.5
Other (Switzerland; Poland; Netherlands; Argentina; Georgia; Montenegro)	6	2.4
Education		
Hauptschule	13	5.5
Realschule	31	13.1
Gymnasium	66	28
University	126	53.4
Profession		
Employed	137	58.1
Unemployed	5	2.1
Student	82	34.7
Pensioner	12	5.1
Summary Statistics		
Variables	Mean	S.D.
Age (18-84)	33.08	12.176

Materials

Instrument

The IPIP Big-Five factor markers consist of a 50-item and 100-item inventory freely available in the public domain (Goldberg, 1999). The current study made use of the 50-item version consisting of 10 items for each of the Big-Five personality factors: Extraversion (E), Agreeableness (A), Conscientiousness (C), Emotional Stability

(ES), Openness (O). For each of the items, which are presented in a short phrase format (e.g., 'Am the life of the party'), 'I' ('Ich' in German) was added at the beginning so that the items would be easier to read. The IPIP items were administered with a 5-point, Likert-type scale ranging from 1 (very inaccurate) to 5 (very accurate) as in the original instrument (Goldberg, 1999).

Translation

Brislin (1986) suggested that when translating existing instruments, researchers should use the translation-back-translation procedure. In accordance with this model, the 50 items of the English language IPIP Big-Five factor markers were first translated independently into German by (1) the present researcher and (2) a graduate of German-English translation studies. The translators' mother tongue is German and they both have a fluent understanding of the English language. Brislin (1986) stated that each target language version of an instrument should then be back-translated to the source language. However, due to limited resources, in this study the translators consensually derived a combined version of the questionnaire.

The translation-back-translation model suggests that the final version of the instrument should be validated by testing for reliability and equivalence using a sample of bilingual subjects. This research did not use bilinguals as recent research has suggested that bilinguals may actually change their personalities slightly as they switch between languages, thereby undermining the accuracy of any findings (Ramírez-Esparza *et al.*, 2006). However, to further test the translation and to make sure the language was appropriate, a German-English bilingual speaker with no knowledge of the Five-factor model compared the original and the German translations. No changes had to be made and the resulting instrument was used for the empirical study (see Appendix 2).

Online Questionnaire Software

The translated version of the aforementioned inventory was transferred to an online questionnaire software (www.smart-survey.co.uk). Participants' responses were automatically exported to an SPSS spreadsheet to be used for further analysis. A further benefit of this method is that unlike the traditional pencil-and-paper format, data collection via the internet allows the order of questionnaire items to be randomized across subjects. Randomization is a valuable tool as it eliminates any order effects that might be present in item responses (Fraley, 2007).

Procedure and Ethical Considerations

Participants were recruited via e-mail and online forums. E-mails were sent to the present researcher's personal contacts (opportunity sample). To help obtain a more heterogeneous sample, a snowball recruitment procedure was used. The initial contacts were requested to forward the email to their contacts and so forth. The email comprised of the researcher's contact details to assure potential participants of the legitimacy of this e-mail and also contained a direct link to the web site www.smart-survey.co.uk. Further, information about the research and the link was published in German speaking forums which permit such requests.

In regards to conducting research via the internet, a number of ethical issues needed to be carefully considered. In traditional pencil-and-paper methods of research

informed consent is obtained by participants signing a consent form. To ensure informed consent potential participants were initially presented with a consent form explaining the purpose of the study, what is expected of them and how the data will be used (see Appendices 3 & 4). They were advised that by clicking the 'I agree' button they are indicating to being 18 years or older and they are consenting to taking part in the research. Compared to traditional ways of obtaining consent, this approach has the advantage of not requiring a signature thereby offering anonymity to participants. Further, participants are informed that they may withdraw from the study at any point for any reason. However, since no identifying information will be requested throughout the research, any information collected up until the point of withdrawal may be used for further analysis.

In order for participants to be debriefed effectively, they were presented with a debrief form after completion of the instrument (see Appendices 5 & 6). The debrief form reiterated the aims of the study and included contact information if participants wished to request further information regarding the research findings. The portal used for the inventory did not allow for computation of the scores therefore not providing participants with immediate feedback about their personalities. However, participants were informed about the procedure to request their individual results on the debrief form. The data collection was automated and the scores were saved to a data base for further analysis.

The present researcher hereby attests that the treatment of human participants conforms to all of the requirements of the British Psychological Society code of ethics (see Appendices 7 & 8).

Results

Methods of sampling adequacy

Of the total sample two participants did not reply to all of the questions. Their scores were therefore excluded from the data analysis resulting in a sample of $N = 234$. Prior to conducting a thorough analysis, two measures of sampling adequacy (MSA) were employed to examine the data in regards to its factorability (see Appendix 9). As recommended by Tabachnick and Fidell (2001), the condition of the distribution of the participants' responses was evaluated through Bartlett's test of sphericity (Bartlett, 1954). Bartlett's test of sphericity was significant ($p < .001$), indicating that the data were adequately distributed to allow an evaluation of the potential factor structure. Next, Kaiser-Meyer-Olkin yielded a value of .82, indicating that the ratio of the number of participants to the number of instrument items was sufficient to run a principal component analysis (PCA). The individual item values were within acceptable limits ranging from .64 to .90, suggesting that the PCA could be conducted without having to remove any unsuitable items.

Factor structure of the 50 IPIP items

Based on the Kaiser criterion (1960), the principal component analysis suggested the extraction of 12 factors with eigenvalues of 1.00 and above (see Appendix 9). However, this method has consistently been found to be inaccurate as the number of eigenvalues greater than one will typically be in the range between one-quarter and

one-third of the number of variables regardless of the actual factor structure (Goldberg & Digman, 1994). The plot of the first 15 eigenvalues is presented in Figure 1. There was no clear discontinuity in the sizes of the eigenvalues in the scree plot and it suggested the extraction of either five, six or seven factors. It was therefore decided to employ parallel analysis which was proposed by Horn (1965) as a means of improving the Kaiser criterion (see Appendix 10). This method has been found to be one of the most accurate methods of factor extraction (Velicer *et al.*, 2000) and suggested in the current data set the extraction of seven factors.

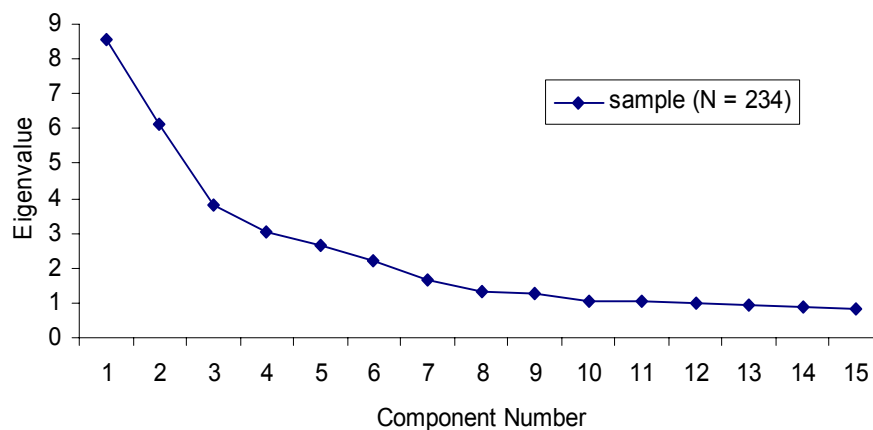


Figure 1: Plot of the eigenvalues on 50-items in the data set

Due to the expected five factor structure of the 50 translated IPIP items, both a five and seven factor solution were compared in regards to the meaningfulness of their factors. It was found that the five factor solution most closely corresponded to the best approximation of simple structure with the fewest number of cross-loadings and yielded the most interpretable solution. The five factor structure was chosen over the seven factor solution as there were no clear conceptual differences between the items representing each of the factors in both solutions. Further, the seven factor solution resulted in a split of the Agreeableness and Openness factor into two.

Five factors accounting for 48.5% of the variance were extracted from the data by PCA and subjected to varimax rotation (see Table 2). Next, all items with weak loadings as well as cross-loadings were discarded. Stevens (1992) recommended that for a sample size of 200 a statistically significant factor loading should be greater than .364. Therefore, all items with a factor loading of less than .364 on a specific component were discarded. One item ('I spend time reflecting on things') did not meet the minimum retaining criteria of .364. Further, items which loaded moderately (at least .35; Clark & Watson, 1995) on two or more factors were also discarded. Two items had cross-loadings on the Extraversion and Openness factors ('I have excellent ideas'; 'I am full of ideas') and one item loaded on Agreeableness and Emotional Stability ('I insult people') and were consequently deleted from the data set. As will be discussed below, these findings may indicate that the concepts related to these items are defined differently in English and German cultures.

Table 2: Five-factor Varimax-Rotated Loadings of the 50 IPIP Items in a German-speaking sample ^a

Item	Component				
	<i>E</i>	<i>A</i>	<i>C</i>	<i>ES</i>	<i>O</i>
Extraversion					
1. I am the life of the party.	<u>.747*</u>	-.006	.022	-.135	.066
6. I don't talk a lot. ^b	<u>.597*</u>	.217	.072	.018	.128
11. I feel comfortable around people.	<u>.559*</u>	.304	.066	.157	-.055
16. I keep in the background. ^b	<u>.706*</u>	.032	-.042	.114	.183
21. I start conversations.	<u>.697*</u>	.260	.037	-.156	.028
26. I have little to say. ^b	<u>.544*</u>	.182	-.015	.104	.324
31. I talk to a lot of different people at parties.	<u>.755*</u>	.052	.048	.041	-.038
36. I don't like to draw attention to myself. ^b	<u>.641*</u>	.017	-.124	.201	.222
41. I don't mind being the centre of attention.	<u>.582*</u>	.036	-.089	.050	.279
46. I am quiet around strangers. ^b	<u>.696*</u>	.072	-.075	.238	-.008
Agreeableness					
2. I feel little concern for others. ^b	-.079	<u>.567*</u>	-.095	.012	.072
7. I am interested in people.	.150	<u>.578*</u>	.070	-.059	.028
12. I insult people. ^b	-.239	<u>.390**</u>	.019	<u>.438**</u>	.001
17. I sympathise with others' feelings.	.079	<u>.795*</u>	.033	-.190	.013
22. I am not interested in other people's problems. ^b	.071	<u>.693*</u>	-.073	.159	.223
27. I have a soft heart.	.245	<u>.698*</u>	.092	.018	.033
32. I do not have a good imagination. ^b	.188	<u>.610*</u>	-.112	.084	.256
37. I take time out for others.	.162	<u>.580*</u>	.137	.018	.149
42. I feel others' emotions. ^{b c}	-.056	<u>-.723*</u>	-.114	.148	-.217
47. I make people feel at ease.	.273	<u>.510*</u>	.011	.046	-.232
Conscientiousness					
3. I am always prepared.	.043	.017	<u>.651*</u>	-.060	.198
8. I leave my belongings around. ^b	-.055	-.120	<u>.600*</u>	.245	-.196
13. I pay attention to details. ^d	-.066	.087	.309	-.132	<u>.418*</u>
18. I make a mess of things. ^b	.208	.068	.131	<u>.524*</u>	.292
23. I get chores done right away.	.056	.073	<u>.710</u>	.099	-.157
28. I often forget to put things back in their proper place. ^b	-.052	-.079	<u>.499</u>	.346	-.014
33. I like order.	.031	-.036	<u>.641</u>	-.070	-.069
38. I shirk my duties. ^b	-.074	.230	<u>.464</u>	.350	-.015
43. I follow a schedule.	.092	.112	<u>.477</u>	.030	.248
48. I am exacting in my work. ^b	.055	-.013	<u>-.622</u>	-.129	-.282
Emotional Stability					
4. I get stressed out easily. ^b	.030	-.151	.001	<u>.763*</u>	.076
9. I am relaxed most of the time.	.260	.026	.130	<u>.575*</u>	-.011
14. I worry about things. ^b	.243	-.175	.007	<u>.692*</u>	.027
19. I seldom feel blue.	.301	-.106	.208	<u>.590*</u>	-.064
24. I am easily disturbed. ^b	.212	-.196	-.048	<u>.694*</u>	.212
29. I get upset easily.	.155	-.123	.058	<u>-.715*</u>	.002
34. I change my mood a lot. ^b	-.164	.101	.084	<u>.771*</u>	.059
39. I have frequent mood swings. ^c	.006	.032	-.146	<u>-.829*</u>	-.072
44. I get irritated easily. ^b	-.093	.117	-.021	<u>.806*</u>	.046
49. I often feel blue. ^c	-.260	-.031	-.187	<u>-.730*</u>	.027
Openness					
5. I have a rich vocabulary.	.279	.035	.135	.050	<u>.584*</u>
10. I have difficulty understanding abstract ideas. ^b	-.091	.111	-.087	.245	<u>.686*</u>
15. I have a vivid imagination.	.243	.289	-.149	-.209	<u>.392*</u>
20. I am not interested in abstract ideas. ^b	-.025	.129	-.133	.202	<u>.633*</u>
25. I have excellent ideas.	<u>.439**</u>	.139	.063	.044	<u>.447**</u>
30. I do not have a good imagination. ^b	.200	.274	-.269	.030	<u>.420*</u>
35. I am quick to understand things.	.193	.103	.149	.104	<u>.573*</u>
40. I use difficult words. ^{b c}	-.251	.083	-.111	.073	<u>-.544*</u>
45. I spend time reflecting on things.	-.223	.263	.310	.061	.304
50. I am full of ideas. ^b	<u>.409**</u>	-.154	-.075	-.039	<u>-.517**</u>

Note. Loadings over .364 are underlined. The highest factor loading for each variable is indicated with an asterisk (*). Cross-loadings are indicated with two asterisks (**). IPIP = International Personality Item Pool; E = Extraversion; A = Agreeableness; C = Conscientiousness; ES = Emotional Stability; O = Openness. The order of the items in the table reflects their order on the IPIP Web site (Goldberg, 2005) ^a *N* = 234. ^b reversed items ^c items deleted due to reliability ^d items deleted due to heterogeneity

In relation to the retained items, all Extraversion items loaded on the same factor, as did all the Agreeableness, Emotional Stability and Openness items. Two of the Conscientiousness items loaded on different factors than expected. The item 'I make a mess of things' loaded most highly on the Emotional Stability factor. It could be argued that someone with low emotional stability might see themselves as a person who makes a mess of things. Therefore this item should be retained but should be reworded for future use in order to amplify the conscientiousness aspect of the item and eliminate the nonspecific aspects that contributed to its loading on the Emotional Stability factor. Further, the item 'I pay attention to details' loaded most highly on the Openness factor. The present researcher was unable to explain the reasons for the item's loading on this specific factor, however, as will be highlighted below, further analysis suggested that this item should be discarded from the item pool.

Homogeneity of the scales

The homogeneity of each of the five scales was assessed using confirmatory factor analysis to test the fit of a single-factor model (Schmitt, 1996; see Appendices 11-15). In relation to dimensionality of the scales, items with loadings less than .4 would be discarded (Brace *et al.*, 2006). For four of the scales all items were retained based on this criterion. The scale measuring the Openness factor included one item ('I pay attention to details') with a loading of .34 and was therefore deleted. It should be noted that this item had previously been identified as not loading on its expected factor (Conscientiousness).

Item Redundancy Analysis

The inter-item correlations for each scale were examined in order to identify any redundant items which do not add unique information to the scale (see Appendices 10-15). The examination of the Extraversion and Conscientiousness scale revealed none of such items; the Agreeableness scale included one redundant item pair ('I feel others' emotions' and 'I sympathise with others' feelings') with an inter-item correlation of .65; as did the Openness scale ('I use difficult words' and 'I have a rich vocabulary') with an inter-item correlation of .54. In relation to the Emotional Stability scale two redundant item pairs were highlighted ('I have frequent mood swings' and 'I change my mood a lot'; 'I often feel blue' and 'I seldom feel blue') with inter-item correlations of .78 and .68 respectively. These findings would suggest that some of the scales could be abbreviated without loss of information (Robins *et al.*, 2001). The items were retained for the next analysis as it was unclear at this point which item of each pair should be discarded.

Internal Reliability of each of the scales

The internal consistency reliabilities were calculated for each of the scales using coefficient alpha (Cronbach, 1951; see Appendix 10-15). The Cronbach's Alpha (R) for Extraversion was high at .84. For Conscientiousness, Agreeableness, Openness and Emotional Stability the initial R values were .65, .69, .52, and .30 respectively. The R values for these scales were considered low; therefore all the items with corrected item total correlation less than .3 were discarded for the Agreeableness, Openness and Emotional Stability scale. When R was calculated on the retained items, R values for Agreeableness, Openness and Emotional Stability increased to

.82, .71 and .76 respectively. It should be noted that all of these items belonged to redundant item pairs highlighted in the previous analysis. In relation to the Conscientiousness scale two items ('I am exacting in my work'; 'I follow a schedule') did not fulfil the above criteria. However, these items were retained as they had not previously been identified during the analyses of heterogeneity and item redundancy of the scale.

Table 3

Internal consistency reliability estimates (coefficient alpha) for the German 50-item IPIP scales in a German-speaking sample^a

Scale	50-Item
	German
Extraversion	.84 ^b (.87)
Agreeableness	.82 ^b (.82)
Conscientiousness	.65 ^b (.79)
Emotional Stability	.76 ^b (.86)
Openness	.71 ^b (.84)

Note. IPIP = International Personality Item Pool; The corresponding values from a U.S. community sample are listed in parenthesis.

^a $N = 234$. ^b after deletion of redundant items

Further, it was felt that the deletion of these items would not significantly impact on the R value of the scale. The Cronbach's Alpha for Conscientiousness therefore remained .65. It can be concluded that each of the five scales of the entire instrument possessed reasonably high construct validity and reliability. The reliability values in this study are comparable to those reported by Goldberg (2005) in an American community sample, although slightly lower for some of the scales. The final R values of each scale are shown in Table 3 and the corresponding U.S. alpha coefficients are given in parentheses.

Discussion

Main Findings and Previous Research

The main purpose of this study was to develop a German version of the IPIP Big-Five factor markers and to investigate the validity of a five-factor structure. The results of the current study provided considerable support for the generalisability of the Five-factor IPIP structure in a German context. The findings demonstrated a substantial congruence between the American and the German IPIP Five-factor structure with only minor deviations from the expected item loadings. The reliabilities of the five factors were high except for the Openness factor which was slightly lower but within acceptable limits. All scales were homogeneous in that items on each scale tap a single factor. These results suggest that the German version of the IPIP Big-Five markers would provide a valuable measure in investigating personality.

The findings of the current study supported the five-factor IPIP structure in a German-speaking sample. The purpose of the International Personality Inventory Pool is an international collaboration within the field of personality research. The freely accessible items allow for comparative and validating studies across nations thereby continuously improving personality assessment. Although the IPIP Big-Five factor markers have been translated into several languages (Goldberg *et al.*, 2005), there is only limited research in relation to the characteristics of these factor markers in other languages and cultures. This research therefore adds to previous studies which have investigated the IPIP Big-Five markers' factor structure in Croatia (Mlačić & Goldberg, 2007) and China (Zheng *et al.*, 2008). These studies also found a clear five-factor structure which was nearly identical to the American structure. The results of previous research and the current study suggest that the IPIP Big-Five factor markers provide a starting point for cross-national comparisons of individual differences.

Limitations

This study attempted to improve on limitations of previous studies in personality research as well as psychology research in general such as translation issues and the samples' limited representation of the population under study. The IPIP Big-Five factor markers were translated in accordance with Brislin's (1986) translation-back-translation procedure. Although thorough discussions between the translators in this study resulted in a final set of items that was considered to be symmetrically translatable to the English counterparts, a number of items needed to be discarded during the analysis stage of this study. This might indicate the presence of item bias which could explain low and cross-factor loadings resulting from mistranslation of certain items. It is therefore recommended that these items should be reworded. In order to validate the final version of the resulting instrument more accurately, a bilingual study to pilot test the translation should be conducted. Dissimilar item statistics between the two versions would indicate poor translations and items could be revised before further analysis.

An alternative explanation might be that some of the variables within this study are not conceptually equivalent. As an example, two of the discarded items ('I have excellent ideas'; 'I am full of ideas') were expected to load on the Openness factor. Whilst in the German version of the IPIP Big-Five factor markers these items did load on this specific factor, they loaded almost equally strong on Extraversion. This would indicate that concepts relating to having ideas are defined differently in English- and German-speaking cultures. Cross-cultural research often suffers from construct bias as a result of an imposed-etic approach in which researchers rely on translations of already existing measures of the construct of interest (Van de Vijver & Leung, 1997). Therefore an emic approach would be favourable in which the instrument is derived locally and captures the indigenous conceptualisation of a construct (Benet-Martínez, 2007). However, this approach is considerably more labour-intensive. The present researcher acknowledges limitations caused by potential item and construct bias but it was felt that this did not compromise the overall factor structure observed.

A further limitation of this study is related to selection bias as participants for this study were not selected using a random sampling technique. Participants were recruited based on the researcher's personal contacts and their contacts. This form

of convenience and snowball sampling has been found to introduce biases that often go unnoticed and are nearly impossible to estimate (Hartmann, 2005). Further, samples recruited by these methods tend to be unrepresentative of the general population (Caspi, 1998). In order to select a more representative sample, participants were also recruited via internet forums. This resulted in a sample which was more diverse than those used in the majority of psychology research which are largely composed of university students (Gosling *et al.*, 2004). However, internet samples are only a modest improvement over traditional sampling methods. For example, the high proportion of participants who have attended university and *Gymnasium* in this study are not representative of the general population. Further, whilst the use of modern technology such as email and internet forums allowed for the recruitment of participants from different locations, it nevertheless limited the sample to those who have access to the internet. A combination of modern and traditional sampling methods may help to overcome these issues and provide a more heterogeneous sample.

Recommendations

Despite the aforementioned limitations, this first attempt to validate the IPIP Big-Five markers in German produced acceptable results. Gow *et al.* (2005) investigated the psychometric properties of the 50-item IPIP Big-Five factor markers and found validating evidence of internal consistency and concurrent validation. Therefore, as a next step, it would be useful to examine the factor structure equivalence of the German instrument developed in this study across different demographic variables such as gender and age. This will contribute to the development of a valid and reliable measurement of personality. In addition, once sufficient reliability and validity data have been collected, it would be useful to examine the external validity of the instrument. As has been discussed above, studies have demonstrated the validity and utility of the FFM in predicting behaviours and outcomes (e.g. Barrick & Mount, 1991; Friedman *et al.*, 1995). It would now be of interest to investigate criterion-related validity of scores on the German version of the IPIP Big-five factor markers. Wiggins (1973, p. 406) argued that prediction “is the sine qua non of personality assessment” and the present research is one step towards this aim.

References

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47 (1, Whole No. 211).
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Bartlett, M. S. (1954). A note on the multiplying factors for various chi-square approximations. *Journal of Royal Statistical Society*, 16, 296-298.

Benet-Martínez, V. (2007). Cross-cultural personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality research* (pp. 170-189). New York, NY: Guilford Press.

Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for Psychologists* (3rd ed.). New York: Palgrave Macmillan.

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural psychology* (pp 137-164). Beverly Hills, CA: Sage.

Brislin, R. W. (1993). *Understanding culture's influence on behavior*. New York: Harcourt Brace.

Caspi, A. (1998). Personality development across the life course. In W. Damon (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* 5th ed., pp. 311-388). New York: Wiley.

Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika*, 12, 197-220.

Cheung, F. M., & Leung, K. (1998). Indigenous personality measures: Chinese examples. *Journal of Cross-Cultural Psychology*, 29, 233-248.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Conn, S. R., & Rieke, M. L. (1994). *The 16PF fifth edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing.

Costa, P. T., & McCrae, R. R. (1989). NEO PI/FFI Manual Supplement. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Ehrhart, K. H., Roesch, S. C., Ehrhart, M., G., & Kilian, B. (2008). A Test of the Factor structure Equivalence of the 50-Item IPIP Five-Factor Model Measure Across Gender and Ethnic Groups. *Journal of Personality Assessment*, 90(5), 507-516.

Fraley, R. C. (2007). Using the Internet for Personality Research: What Can Be Done, How to Do It, and Some Concerns. In R. W. Robins, R. C. Fraley & R.F.

Krueger (Eds.). *Handbook of Research Methods in Personality Psychology* (pp. 130-148). New York: The Guilford Press.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., & Tomlinson-Keasey, C. (1995). Psychosocial and behavioral predictors of longevity: The aging and death of the "Termites". *American Psychologist*, *50*, 69-78.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26-42.

Goldberg, L. R. (1995). What the hell took so long? Donald Fiske and the Big-Five factor structure. In P. E. Shrout & S. T. Fiske (Eds.). *Personality research, methods, and theory: Festschrift honouring Donald W Fiske* (pp. 29-43). Hillsdale, NJ: Erlbaum.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Marveled, I. J. Deary, F. de Fruyt, & F. Ostendorf (Eds.). *Personality psychology in Europe* (Vol. 7, pp. 7-28). Tilburg: Tilburg University Press.

Goldberg, L. R. (2005). *International Personality Item Pool: A scientific collaboratory for the development of advanced measures of personality traits and other individual differences*. Retrieved February 28, 2010, from the International Personality Item Pool Web site: <http://ipip.ori.org>

Goldberg, L. R., & Digman, J. M. (1994). Revealing structure in the data: Principles of exploratory factor analysis. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 216-242). New York: Springer.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, G. (2005). The international personality item pool and the future and public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O.P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, *59*, 93-104.

Gough, H. G., & Bradley, P. (1996). *CPI manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's "IPIP" Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, *39*, 317-329.

Hartmann, D. P. (2005). Assessing growth in longitudinal investigations: Selected measurement and design issues. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 319-339). Malden, MA: Blackwell.

Hendriks, A. A. J., Hofstee, W.K.D., & De Raad, B. (1998). Short behaviour-descriptive sentences as units of personality measurement. In J. Bermudez, B. De

Raad, J. De Vries, A. M. Perez-Garcia, A. Sánchez-Elvira, & G. L. Van Heck (Eds.). *Personality psychology in Europe* (Vol. 6, pp. 40-50). Tilburg, The Netherlands: Tilburg University Press.

Hogan, R., & Hogan, J. (1992). *Hogan personality inventory manual*. Tulsa, OK: Hogan Assessment Systems.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.

John, O. P. (1989). Towards a taxonomy of personality descriptors. In D. M. Buss & N. Cantor (Eds.). *Personality psychology: Recent trends and emerging directions* (pp. 261-271). New York: Springer-Verlag.

John, O., Donahue, E., & Kentle, R. (1991). The "Big Five" inventory – versions 4a and 54. *Technical Report*, Institute of Personality Assessment and Research, Berkeley, CA: University of California, Berkeley.

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* (2nd ed., pp. 102-138). New York: Guilford.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Kline, P. (1994). *An Easy Guide to Factor Analysis*. New York: Routledge.

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin*, 120, 323-337.

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparison. In R. R. McCrae & J. Allik (Eds.). *The Five-factor model of personality across cultures* (pp. 105-125). New York: Kluwer Academic/Plenum.

Mlačić, B., & Goldberg, L. R. (2007). An analysis of a cross-cultural personality inventory: The IPIP Big-Five factor markers in Croatia. *Journal of Personality Assessment*, 88, 168-177.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks, CA: Sage Publications, Inc.

Ramírez-Esparza, N., Gosling, S. D., Benet-Martínez, V., Potter, J. P., & Pennebaker, J.W. (2006). Do bilinguals have two personalities? A special case of cultural frame-switching. *Journal of Research in Personality, 40*, 99-120.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353.

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martinez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and Profiles of Human Self-Description Across 56 Nations. *Journal of Cross-Cultural Psychology, 38*, 2, 173-212.

Stevens, J. P. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. Technical Report, USAF, Lackland Air Force Base, TX.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. London: Sage.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.). *Problems and solutions in human assessment: Honoring Douglas Jackson at seventy* (pp. 41-71). Boston, MA: Kluwer.

Zheng, L., Goldberg, L. R., Zheng, Y., Zhao, Y., Tang, Y., & Liu, L. (2008). Reliability and concurrent validation of the IPIP Big-Five factor markers in China: Consistencies in factor structure between internet-obtained heterosexual and homosexual samples. *Personality and Individual Differences, 45*, 649-654.