UNIVERSITY OF CAMBRIDGE

Faculty of Politics, Psychology, Sociology and International Studies

Michał Kosiński

# Protocol validity indices in the sample from an on-line personality questionnaire.

Cambridge 2009



michal@neworder.pl

TABLE OF CONTENTS:

## Introduction

On-line implementations of the paper-and-pencil personality measures allow researchers to collect huge samples of personality scores inexpensively and in a way that is convenient for both sides – administrators and participants (Buchanan, Johnson, & Goldberg, 2005; Johnson, 2005). An on-line questionnaire can be completed at the time and place chosen by the respondent rather than in a less familiar setting (e.g. laboratory) created by the test administrator, which is not only convenient but also increases the ecological validity of the results (Reips, 2000; Buchanan, 2002). Moreover, it was found that individuals are more candid in online questionnaires than in those administered in face-to-face conditions, especially when the participation is anonymous (Gosling, Vazire, Srivastava, & John, 2004). Finally, the assessment software can automatically collect and process individual records providing instant feedback to the participants and administrators. Due to the above advantages, there is a growing interest in the web-mediated personality assessment and several well established paper-and-pencil questionnaires have been used in the on-line settings (e.g. Buchanan, Johnson, & Goldberg, 2005; Johnson, 2005; Soto, John, Gosling, & Potter, 2008). The web-mediated personality instruments are not only used for research purposes, but were also implemented in such areas as marketing, commercial and educational assessment (Buckley & Williams, 2002) and tele-health services (Buchanan, 2002).

On the other hand, the web-mediated assessment poses several serious challenges to the administrators. For instance, although ecological validity increases when the participants can choose the place and time to undertake the test, it also means that the testing situation is neither standardized nor controlled by the administrator (Reips, 2000; Buchanan & Smith, 1999). Moreover, web questionnaires may be affected by raised numbers of unreliable and not motivated participants, which would response in a careless, dishonest, or mischievous way

(Buchanan & Smith, 1999; Reips, 2000). Finally, some results suggest that web-questionnaire findings may be inconsistent with findings from traditional methods (Gosling, Vazire, Srivastava, & John, 2004).

The above facts imply that even if the online measure is a direct copy of an established paper-and-pencil personality questionnaire no real weight can be attached to its results until its psychometric properties are verified: *"...it is clear that one cannot simply mount an existing questionnaire on the world wide web and assume that it will be exactly the same instrument* (Buchanan, Johnson, & Goldberg, 2005, page 125)*."* Before the psychometric properties of any personality questionnaire can be assessed correctly one should discard the individual protocols that are the product of inattentive, linguistically incompetent, deliberately manipulating or randomly answering participant.

## Protocol Validity in on-line instruments

*Protocol validity* refers to whether an individual record can be scored with the standard scoring schema (Kurtz & Parrish, 2001). Even a well-established and validated personality measure can produce invalid results in individual cases due to the *linguistic incompetence, careless inattentiveness, and deliberate misrepresentation* (Johnson, 2005). As it will be shown below, the problem of protocol invalidity is particularly relevant to the on-line personality assessment.

### *Linguistic incompetence*

Web-based personality tests can be easily administered to participants from different backgrounds worldwide. It is a great advantage, but it also leads to the elevated risk of invalidating the protocol due to the participant's linguistic incompetence. Non-native speakers or those who use different variety of the test's language may misunderstand a question (Johnson, 2005). However, even those with proper language competences can interpret an

item in an idiosyncratic way, which would also lead to the diminished protocol validity (Johnson, 2002).

***Carelessness and inattentiveness***

Invalid protocols may also be produced by the inattentive and careless participants who skip or misread the test items, answer in the wrong areas of the answer sheet or respond randomly (Kurtz & Parrish, 2001). Tests administered on the Internet have several features that may lead to higher levels of inattentiveness and carelessness. Firstly, the ease of responding and instant feedback may rush the respondents to quickly submit their protocol without paying as much attention to their answers as in the paper-and-pencil measures (Johnson, 2005). Secondly, the administrator has little or no control over the circumstances in which the test is taken, so it is possible that a participant is simultaneously engaged in other activities (e.g. instant messaging). Thirdly, the psychological distance between the administrator and the participants, caused by the lack of face-to-face contact, may decrease the feeling of accountability in respondents, especially if the questionnaire is anonymous (Gosling, Vazire, Srivastava, & John, 2004).

***Misrepresentation***

A third threat to the protocol validity is any deliberate and conscious attempt to manipulate one's personality score. The numbers of the protocols invalidated by the deliberate misrepresentation may be elevated in the web-questionnaires, since the lack of personal contact, the psychological distance and feeling of anonymity in web-mediated communication may encourage individuals to use misrepresentation strategies or even construct completely fake identities (Caspi & Gorsky, 2006). There are two main forms of misrepresentation, "faking good" and "faking bad" (Kurtz & Parrish, 2001). Respondents that use the first of those misrepresentation strategies claim to be better adjusted, more attractive and competent

than they are perceived in everyday life. On the opposite, “faking bad” respondents try to appear more incompetent or maladjusted than they really are.

## Protocol validation methodology

The most direct way of validating the protocol would be to compare the personality scores measured by the tested instrument with another, confident source of information about respondent’s personality (Johnson, 2005). The most popular sources of such validating information are: ratings from respondent’s peers, self reports of behaviour, life events associated with personality or retesting respondent with another, established personality measure (Buchanan, Johnson, & Goldberg, 2005). Nonetheless, this study examines the protocol validation methods that rely solely on the internal data properties, such as: missing responses, response strings, response patterns and internal consistency.

### *Missing responses*

The individual protocols beset with numerous missing responses have reduced accuracy, even if the remaining responses are product of attentive and honest individual. However, in many cases the items are left blank due to respondent’s linguistic incompetence, carelessness or inattentiveness. Therefore, removing the records with particularly high numbers of missing responses is one of the most outright ways of discarding invalid protocols and improving the significance and the accuracy of the research results (Johnson, 2005).

Detecting the protocols containg missing responses is simple, unlike deciding how many items may be left blank before an individual protocol is considered invalid. Johnson (2005) reported the average of 1.2% missing responses in 300-item on-line version of IPIP proxy for NEO-PI-R and the average of .1% - .5% in several paper-and-pencil samples. He proposed to discard an individual protocol from the on-line version of IPIP instrument if it has more than 10 missing responses, which eliminated 2.9% of his sample.

***Random responding***

Random responding is defined as a strategy in which responses are made without regard to item content (Kurtz & Parrish, 2001). Respondents may engage in deliberate random responding, but it is usually caused by individual's inattentiveness or linguistic incompetence (Johnson, 2005). Obviously, the protocols that are partially or wholly produced by random responding have reduced or no validity at all (Costa & McCrae, 1997). Records invalidated by random responding may consist of long *consecutive identical response strings, response patterns* and are characterized by *low internal consistency*.

*Long consecutive identical responses strings and response patterns*

Respondent who wants to finish the questionnaire as quickly as possible, is too tired, distracted or linguistically incompetent to read and understand a question may simply continually mark the same answer category in subsequent items or answer in patterns. Response patterns are strings composed of the recurring sets of 2 or more consecutive responses and consist at least two different response categories. For example the string: 3, 4, 3, 4, 3 is a 5 item long response pattern composed of 2 item long subsets (3, 4), while the string: 1, 1, 2, 1, 1, 2 is a response pattern composed of 3 item long subsets (1, 1, 2) and 6 item long.

Detecting consecutive identical response strings (CIRS) and response patterns is relatively easy, however it is problematic to decide what lenght of such string should invalidate the protocol. Costa and McCrae (2005) reported that in a sample of highly cooperative and attentive 983 volunteers filling the paper-and-pencil 240 item NEO-PI-R questionnaire no one used the same response for more than 6, 9, 10, 14, 9 consecutive times ("strongly disagree", "disagree", "neutral", "agree", "strongly agree" respectively). They suggested that NEO-PI-R instruments that contain longer strings of the same responses should be considered as possibly invalid due to inattentive responding. Johnson (2005) applied scree-

like test (Cattell, 1966) to the CIRS in the sample from 300-item on-line version of IPIP proxy for NEO-PI-R. He detected sudden drops in the frequency of the CIRS lengths for the following values: 9, 9, 8, 11 and 9 (1 to 5 on a Likert response scale, respectively). Nevertheless, he decided to use the more conservative values suggested by Costa and McCrae (2005). Although, there was no literature regarding the response patterns available to the author of this essay, it seems that the methods and cut-off points designed for CIRS may be effectively applied also to the response pattern strings.

*Internal consistency*

Random responding not necessarily entails answering in patterns or CIRS. However, as random responding is by definition content independent (Kurtz & Parrish, 2001) it decreases the internal response consistency, and therefore may be detected by means of item inconsistency measures. There are several approaches to measuring response consistency in personality questionnaires. In the semantic antonym method (Goldberg & Kilkowski, 1985) respondent's answers in items that are semantically opposite (e.g.: "Dislike myself" and "Am very pleased with myself") are compared, and if they are not contrasting it is considered as an index of inconsistency. Another similar approach uses pairs of items that have the highest negative correlation, called psychometric antonyms. Psychometric antonyms are not necessarily semantic antonyms and should not belong to the same scale. As in the semantic antonym approach, consistent responders are expected to answer the psychometric antonyms in opposite directions. Johnson (2005) proposed to discard protocols as invalid when the Goldberg's Psychometric Antonym Coefficient (GPA) is lower than .03.

Another inconsistency measurement method, called Jackson's Individual Reliability Coefficient (JIR), was proposed by Douglas Jackson (SIGMA Assessment Systems, Inc., 2008). To calculate JIR, items that belong to one scale are numbered sequentially in order of appearance and split in odd and even-numbered subsets. Scores are calculated for those

subsets and odd- and even-numbered half-scale scores are correlated and corrected for decreased length by the Spearman-Brown formula. Jackson (1977) proposed that individuals that produced an individual reliability coefficient of less than 0.3 can be categorized as inattentive, careless, uncooperative or linguistically incompetent. This cut-off point was later endorsed by Johnsons (2005) in his online IPIP sample.

This time again the decision of how inconsistent a protocol must be to be rejected is highly problematic. Moreover, some findings suggest that level of consistency is an individual difference and inconsistent protocols may be a product of conscious and attentive responding. Costa and McCrae (1997) compared 841 NEO-PI-R domain scores with Goldberg's Big Five (1992) adjective markers (gathered 7 months earlier) in groups of high, medium and low level of consistency. Out of five personality scores, only agreeableness exhibited an appreciable reduction in convergent validity on greater levels of response inconsistency (Costa & McCrae, 1997). Those results were also replicated in Kurz and Parrish (2001) study. Costa and McCrae (1997) as well as Kurz and Parrish (2001) concluded that the level of internal consistency is more of an individual difference than index of protocol validity. These conclusions were supported by Johnson (2005), who found that GPA correlates with Neuroticism and Openness and has a distribution close to normal (which is characteristic for individual differences). Moreover, in his sample the highly consistent records did not produce a clearer factor structure than those of low consistency.

***Misrepresentations***

The accuracy of personality ratings can be seriously distorted by respondents that attempt to manipulate their score. While it is difficult to recognize individuals creating completely fake personality profiles relying solely on the internal item level data properties, it is possible to detect those who try to appear better or worse than they really are.

The uncommon virtues approach allows to detect those who exaggerate in their 'faking good' attempts and claim to posses uncommon virtues (e.g. VIRTUES scale in Multidimensional Personality Questionnaire; Piedmont, McCrae, Riemann, & Angleitner, 2000). In another strategy the responses on the pairs of items that are highly socially desirable but opposite in content are compared, and those who answer 'true' to both are considered more focused on the social desirability than item content (e.g. the Desirable Response Inconsistency scale designed for NEO-PI-R questionnaire; Piedmont, McCrae, Riemann, & Angleitner, 2000). Schinka et al (1997) proposed two scales composed of NEO-PI-R items that are designed to identify respondents that attempt to present themselves as exceptionally good or exceptionally bad (Positive Presentation Management (PPM) and Negative Presentation Management (NPM) scales)

Another method of detecting the respondents that systematically bias their responses in attempt to misrepresent themselves was proposed on the International Personality Item Pool website (2009). It is called the Social Desirability Coefficient (SDC) and is calculated by correlating each item in a questionnaire with the average score for that item across the whole sample. High, low or close to zero correlations may be considered as a sign of faking good, faking bad or responding randomly.

## Method

### *The goal of the study*

The goal of this study is to explore the possibilities of assessing the individual protocol validity relying solely on the internal item level data properties in the sample from the online version of 100-item IPIP Five Factor Model Questionnaire.

***Data acquisition environment***

Sample used in this research was acquired from the database of Mypersonality.org. Mypersonality is a Facebook application that offers its users free personality assessment together with some extra features like comparisons with friend's personality profiles. More details regarding Mypersonality application and its testing procedures can be found in the Appendix 2.

***Personality measure***

Mypersonality.org uses the 100-item IPIP representation of the domain construct of the Costa and McCrae's Fife Factor Model employed in NEO-PI-R (Costa & McCrae, 1992). IPIP scales correlate highly with the corresponding NEO-PI-R domain scores with the correlation coefficients ranging from 0.88 to 0.93 (IPIP.org, 2009). IPIP proxy for NEO-PI-R domain scales are widely used in online personality assessment and proved to be useful and reliable in the research (Buchanan, Johnson, & Goldberg, 2005). IPIP domain scales have been shown to outperform the matching NEO-PI-R constructs as predictors of a number of self-reported behavioural indices (Goldberg, et al., 2006).

***Participants***

The sample used in this research was received from Mypersonality.org on the 9th of January 2009. It consisted 182,922 individual protocols composed of the user id and item level scores of the 100 item personality questionnaire. The preliminary analyses of the missing answers frequency showed that there is a steep rise in numbers of cases with 10, 20, 30, 40, 50, 60, 70, 80 and 90 missing answers. The closer investigation showed that in most cases missing answers occur in 10 item strings and start from item number 11, 21, 31 and so on. The problem was reported to Mypersonality and they confirmed that in some cases blocks of ten answers are lost. The protocols that were invalidated by this problem were removed.

The number of records in the sample was reduced by 27.420 records (from 210.342 to 182.922) before proceeding to further analyses.

## *Analyses*

### *Consecutive identical responses and response patterns*

The SPSS script was written to measure the maximum length of the Consecutive Identical Response strings of each response category for each individual protocol. Another script scanned the item level data for the response strings composed of 2 to 7 items long subsets. To establish the cut-off points appropriate for the instrument under examination, maximum CIRS and response patterns frequencies were analysed with scree-like test and compared with the figures suggested by Costa and McCrae (2005).

### *Missing responses*

Missing responses frequency curve for the current sample was compared with the values found in Johnson's (2005) paper to recommend the maximum missing responses string length for the Mypersonality sample.

### *Internal consistency*

Two consistency measures were calculated in the current study: Goldberg's Psychometric Antonym and Jackson's Individual Reliability coefficients.

The quality of the factor structure expressed by total variance explained, Kaiser-Meyer-Olkin Measure of Sampling Adequacy and factor loadings were compared on different levels of internal consistency measures to check if protocols with higher internal consistency produce a clearer factor structure. To check if response consistency could be regarded as an individual difference, GPA and JIR coefficients were correlated with personality scores and their distribution curves were examined. Finally, both internal consistency measures were entered into the Factor Analysis to check if they load on any latent factor.

*Misrepresentation*

The Social Desirability Coefficient (IPIP.org, 2009) was used to measure the amount of misrepresentation in individual protocols. The distribution of the scores was analysed to propose norms for discarding manipulated responses. The factor structure was analysed on different levels of SDC, to check if there is a relation between the quality of the factor structure and the level of misrepresentation.

## Results

### *Consecutive identical responses*

The means and standard deviations of the longest consecutive identical response strings together with numbers of observations are presented in Table 1. Respondents rarely used response strings longer than two and the average maximum CIRS in the individual record varied from 2.08 (SD=1.55) for answer category 1 ('Very inaccurate') to 3.1 (SD=1.28) for answer category 4 ('Moderately accurate').

Table 1. Frequencies of the maximum consecutive identical response strings of each response category

| | Response category | | | | |
|---|---|---|---|---|---|
| | 1 (Very inaccurate) | 2 (Moderately inaccurate) | 3 (Neither) | 4 (Moderately accurate) | 5 (Very accurate) |
| Mean | 2,08 | 2,50 | 2,43 | 3,10 | 2,39 |
| SD | 1,55 | 0,98 | 1,83 | 1,28 | 2,19 |
| Longest CIRS 0 | 1638 | 950 | 1501 | 703 | 1273 |
| 1 | 54814 | 19019 | 32914 | 6930 | 37835 |
| 2 | 71905 | 79051 | 79593 | 47773 | 71864 |
| 3 | 43164 | 61568 | 43817 | 74564 | 52242 |
| 4 | 8419 | 17812 | 16297 | 34362 | 12551 |
| 5 | 2267 | 3354 | 5286 | 11880 | 4252 |
| 6 | **417** | 831 | 1939 | 4212 | 1357 |
| 7 | 156 | 241 | 780 | 1645 | 933 |
| 8 | 42 | 63 | 359 | 539 | 268 |
| 9 | 20 | **16** | 142 | 175 | **111** |
| 10 | 17 | 6 | **77** | 64 | 44 |
| 11 | 7 | 3 | 51 | 23 | 43 |
| 12 | 5 | 0 | 31 | 10 | 16 |
| 13 | 3 | 0 | 20 | 9 | 9 |
| 14 | 3 | 1 | 9 | **7** | 6 |
| 15 | 2 | 1 | 7 | 3 | 7 |
| 16 | 1 | 0 | 8 | 0 | 3 |
| >16 | 42 | 6,0 | 91 | 23 | 108 |
| Sum: | 182922 | 182922 | 182922 | 182922 | 182922 |

*Note:* The maximum response lengths found by Costa and McCrae (1997) and endorsed by Johnson (2005) are in boldface. The maximum response lengths suggested for the instrument under examination are followed by a double horizontal line.

***Response patterns***

The frequencies of the response patterns that were found in individual protocols are presented in Table 2.

Table 2 The frequencies of the response patterns composed of 2 to 7 item long subsets

| | Response pattern interval | | | | | |
|---|---|---|---|---|---|---|
| | 2 (ABAB) | 3 (ABCABC) | 4 (ABCD ABCD) | 5 (ABCDE ABCDE) | 6 (ABCDEF ABCDEF) | 7 (ABCDEFG ABCDEFG) |
| Mean | 4,65 | 5,50 | 6,81 | 7,91 | 8,77 | 9,55 |
| SD | 1,16 | 0,92 | 1,23 | 1,11 | 1,17 | 0,97 |
| 0 | 109 | 115 | 107 | 112 | 4841 | 113 |
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | 12084 | | | | | |
| 4 | 80120 | 12177 | | | | |
| 5 | 61377 | 96043 | 4701 | | | |
| 6 | 22201 | 52366 | 75652 | 5279 | | |
| 7 | 4855 | 16285 | 67969 | 66346 | 79678 | |
| 8 | 1407 | 4535 | 23219 | 67979 | 64300 | 12209 |
| 9 | 598 | 972 | 8326 | 32248 | 23766 | 89348 |
| 10 | 91 | 286 | 2080 | 6495 | 7739 | 57541 |
| 11 | 42 | 99 | 579 | 2561 | 1612 | 17259 |
| 12 | 9 | 25 | 146 | 1070 | 520 | 4623 |
| 13 | 9 | 9 | 77 | 423 | 210 | 1344 |
| 14 | 2 | 4 | 26 | 223 | 76 | 331 |
| 15 | 4 | 5 | 13 | 90 | 29 | 94 |
| 16 | 0 | 1 | 4 | 37 | 18 | 33 |
| 17 | 1 | 0 | 3 | 23 | 8 | 12 |
| 18 | 0 | 0 | 1 | 12 | 2 | 7 |
| 19 | 1 | 0 | 2 | 13 | 3 | 2 |
| >19 | 12 | 0 | 17 | 11 | 14 | 6 |
| Sum: | 182922 | 182922 | 182922 | 182922 | 182922 | 182922 |

Maximum length of a response pattern

Note: The cut off points on the response pattern lengths that were recommended in the current study are followed by a double horizontal line.

***Missing responses***

The frequencies of the missing responses are shown in Table 3. The average number of missing responses in the sample was 0.34 (SD= 1.23). The number of missing responses in the individual protocol did not correlate with any of the personality scores.

Table 3. Frequencies of missing responses

| Number of missing responses | Frequency | Percent | Cumulative Percent |
|---|---|---|---|
| 0 | 143387 | 78,4 | 78,4 |
| 1 | 29084 | 15,9 | 94,3 |
| 2 | 6678 | 3,7 | 97,9 |
| 3 | 1938 | 1,1 | 99,0 |
| 4 | 741 | 0,4 | 99,4 |
| 5 | 338 | 0,2 | 99,6 |
| 6 | 177 | 0,1 | 99,7 |
| 7 | 112 | 0,1 | 99,7 |
| 8 | 72 | 0 | 99,8 |
| 9 | 50 | 0 | 99,8 |
| 10 | 45 | 0 | 99,8 |
| 11 | 29 | 0 | 99,9 |
| 12 | 26 | 0 | 99,9 |
| 13 | 18 | 0 | 99,9 |
| 14 | 20 | 0 | 99,9 |
| 15 | 17 | 0 | 99,9 |
| 16 | 20 | 0 | 99,9 |
| >17 | 170 | 0,1 | 100,0 |
| Sum: | 182922 | 100 | |

***Internal consistency***

The inter-item correlations for 100 test items were analysed to find the 15 psychometric antonyms that are listed in Appendix 1. Due to the small number of items in each of the scales the rules of choosing the psychometric antonyms were relaxed, and the antonyms coming from the same scale were accepted. The average value of Goldberg's psychometric antonym coefficient in this sample was 0.61 (SD=.26). The average JIR equalled 0.73 (SD=.38). The frequency graphs of JIR and GPA are presented in Figure 1 and Figure 2. The skewed shape of both frequency graphs was as expected when most participants provide consistent answers. The correlations between internal consistency measures and personality scores are shown in Table 4.

Table 4. Correlations between internal consistency measures and personality scores

| | Personality Score | | | | | |
|---|---|---|---|---|---|---|
| | O | C | E | A | N | GPA |
| GPA | 0.208 | 0.170 | 0.325 | 0.108 | -0.224 | 1 |
| JIR | 0.227 | 0.111 | 0.044 | 0.215 | -0.302 | 0.306 |

Note: All of the correlation coefficients presented above are significant at .001 level

Sample was split into tertiles on each measure of consistency, and the Varimax Factor Analysis with 5 latent factors was run on the data in each subset. The Total Variance Explained, Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) and factor loadings were compared between those subsets. The results are shown in Table 5. All above values demonstrating goodness of factor model fit, increased concurrently with the values of consistency measures. The Varimax factor analysis results showed that GPA and JIR coefficients did not significantly load on any of the latent factors.

Table 5. Comparison of Varimax Factor Analysis between tertiles on Goldberg's (GPA) and Jackson's (JIR) internal consistency measures

| | | Factor Analysis | | | | | |
|---|---|---|---|---|---|---|---|
| | | Total Variance Explained | | Kaiser-Meyer-Olkin Measure | | Loading of item 1 on Openness scale | |
| | | GPA | JIR | GPA | JIR | GPA | JIR |
| Tertil | Low | .28 | .26 | .93 | .92 | .42 | .41 |
| | Medium | .35 | .36 | .95 | .96 | .44 | .44 |
| | High | .42 | .43 | .97 | .98 | .43 | .46 |

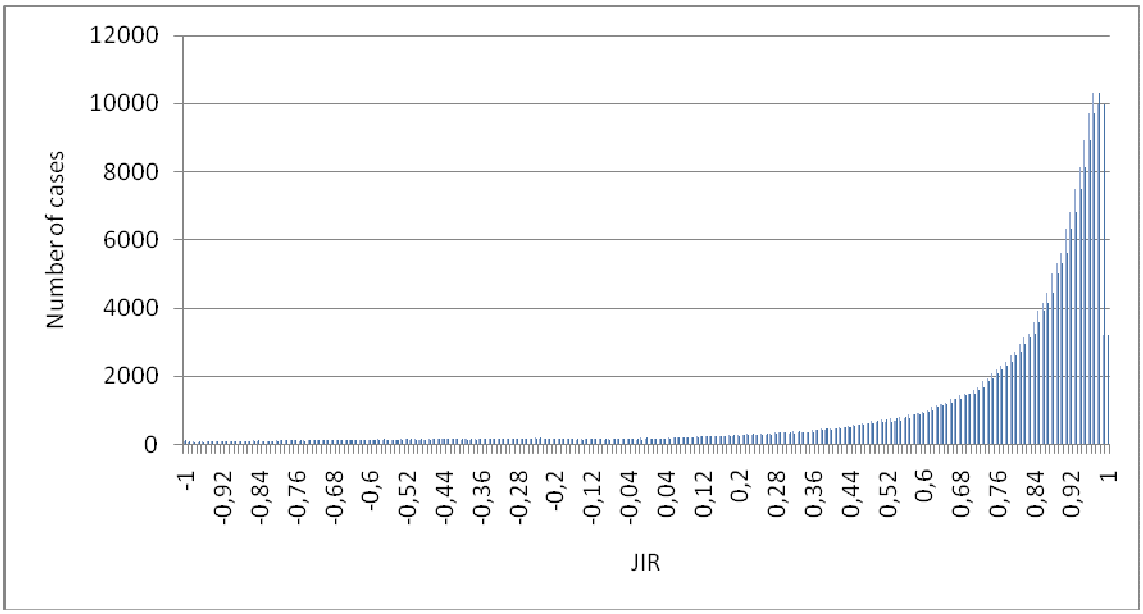



Figure 1. Frequency graph for the Jackson Individual Reliability (JIR) coefficient

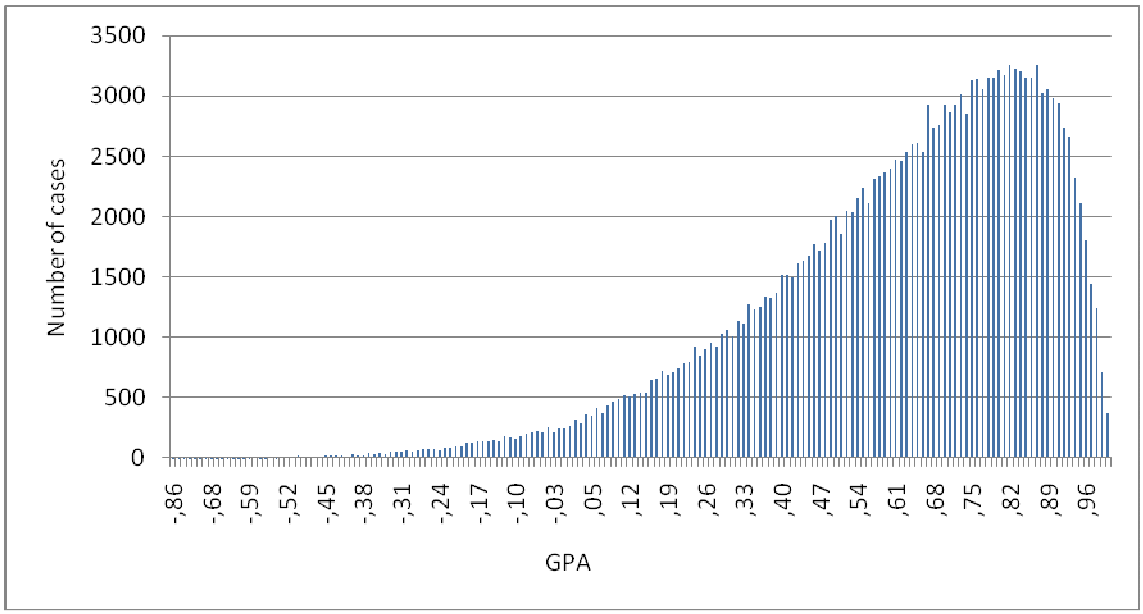



Figure 2. Frequency graph for the Goldberg's psychometric antonym coefficient (GPA)

***Misrepresentation***

The average social desirability coefficient in the current sample equalled 0.48 (SD=.23). The frequency curve of this measure is shown in Figure 3. The sample was split into deciles on the SDC measure. The Varimax Factor Analysis with 5 latent factors was run on each subset. The Total Variance Explained, Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO), factor loadings and the average reliability of 5 personality scales were compared between those subsets. The results are shown in Table 6. There was a steady decline in the factor structure quality with the increase of the SDC level.

There were considerable correlations between SDC and personality scores that vary from 0.66 (Extraversion) to -0.63 (Neuroticism) (Table 7).

Table 6. Comparison of Varimax Factor Analysis and scale alpha reliability between deciles on the Social Desirability Coefficient

| Decil | Cut-off point | Factor Analysis | | Scales reliability | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Total Variance Explained | Kaiser-Meyer-Olkin Measure | O | C | E | A | N | MEAN |
| 1 | | 32 | .94 | .83 | .90 | .89 | .86 | .88 | .87 |
| 2 | .16 | 32 | .94 | .82 | .90 | .88 | .85 | .88 | .86 |
| 3 | .30 | 31 | .94 | .81 | .89 | .88 | .84 | .88 | .86 |
| 4 | .39 | 30 | .93 | .81 | .89 | .88 | .83 | .87 | .86 |
| 5 | .47 | 29 | .93 | .79 | .89 | .87 | .83 | .87 | .85 |
| 6 | .52 | 27 | .92 | .78 | .88 | .87 | .82 | .86 | .84 |
| 7 | .58 | 26 | .92 | .76 | .88 | .86 | .81 | .86 | .83 |
| 8 | .63 | 25 | .91 | .74 | .87 | .85 | .80 | .85 | .82 |
| 9 | .67 | 24 | .90 | .72 | .87 | .85 | .79 | .85 | .82 |
| 10 | .72 | 24 | .92 | .73 | .86 | .84 | .78 | .84 | .81 |

Table 7. Correlations between Social Desirability Coefficient and personality scores

| | Personality Measure | | | | |
|---|---|---|---|---|---|
| | O | C | E | A | N |
| SDC | 0.55 | 0.52 | 0.66 | 0.53 | -0.63 |

Note: All of the correlation coefficients presented above are significant at the 0,001 level

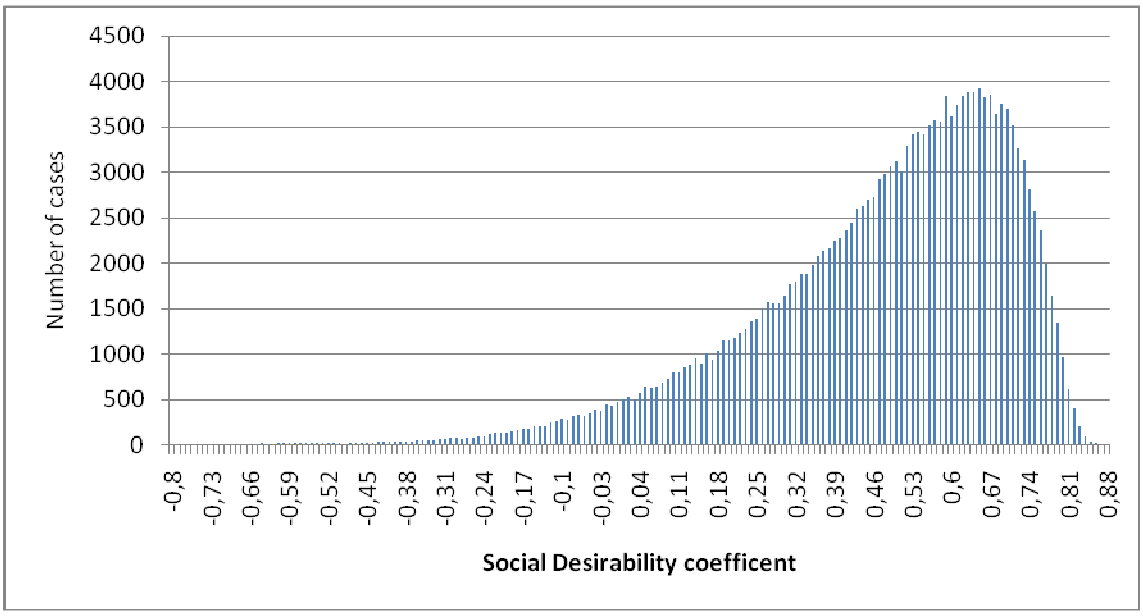



Figure 3. Frequency graph for the Social Desirability Coefficient

## Discussion

The present study investigated the protocol validity indices in item level data of a relatively big online sample from 100 items IPIP personality questionnaire. Results presented above allow proposing norms necessary to exclude invalid protocols in the samples that will be acquired from the same or similar sources in the future.

### *Consecutive identical responses and response patterns*

When a respondent uses the same response category (e.g. “Moderately Accurate”) or set of responses (e.g. 4, 5, 4, 5, 4, 5) repeatedly for all items in the questionnaire, he or she is apparently not giving sincere answers that could produce a valid personality score. However, such extreme situations are rare (in the current sample less than 5 permilles) and more often participants lose their focus for the part of the inventory, especially towards the end of the test (Morey & Hopwood, 2004).

It is difficult to judge when a consecutive identical response string becomes too long to be a product of attentive and honest responding. Applying to current sample the maximum CIRS length suggested by Costa and McCrae (2005) and endorsed by Johnson (2005) would remove 601 records (.3% of all), which is much less than the 6.3% records removed by the same maximum CIRS values in the online sample examined by Johnson (2005). This may be explained by the fact that current instrument was nearly 3 times shorter than the one used by Johnson (2005, 240 items) and Costa and McCrae (1997, 300 items), hence respondents were less likely to get tired or lose their focus. As a consequence it was decided that the norms suggested by Costa and McCrae (2005) should be tightened, also because the smaller number of test items in the current instrument made its scales more prone to the invalidation due to random responding.

The cut-off points for the current sample were estimated using a scree-like-test (Cattel, 1966) suggested by Johnson (2005). An examination of the frequency curve showed an 'elbow' after 6, 6, 8, 8, and 7 consecutive responses (for response categories 1, 2, 3, 4, 5 respectively). There were 1,911 (1% of total) individual records that fell over this limit in the sample under examination.

As it was mentioned before there was no literature on the subject of response patterns available to the author. The analysis of the scree-plot[1] showed a change in the slope of the curve after response patterns that were 9, 10, 11, 12, 13 and 14 items long (respectively for patterns composed of subsets containing 2, 3, 4, 5, 6 and 7 items). Cut-off points proposed above would lead to the removal of 1,793 records (1% of total) in the current sample.

***Missing responses***

While a certain number of the items may be left blank by accident, large amounts of missing responses indicate inattentiveness or carelessness of the respondent. The average number of missing responses in the sample was .34% (out of 100 items). This value is comparable to the .1% - .5% missing responses in some paper-and-pencil inventories (Johnson, 2005). Surprisingly, this figure is nearly four times smaller than 1.2% missing responses in the comparable study on the on-line data (Johnson, 2005), which may suggest that, at least in the present on-line instrument, the participants were no less attentive than in paper-and-pencil questionnaires.

Johnson (2005) set a cut-off point on 10 responses in the 300 item questionnaire which removed 2.9% protocols in his sample. Again, as the current instrument was much shorter, the decision was made to tighten the limits. It is possible that this produced some false positives, however the size of the sample allowed to devote some cases to ascertain the highest possible level of protocol validity. The analysis of the scree-plot showed that the slope flattened after 3

[1] Alternatively, the probability theory or pseudo random cases analysis may be used

missing responses, so this value was selected as a cut-off point. There are 1,835 records (1% of the sample) that have more missing responses than 3.

***Internal consistency***

The nearly bell-shaped distribution of the GPA frequency curve and considerable correlations between internal consistency measures and personality traits (nearly 0.33 between GPA and Extraversion and -0.30 between JIR and Neuroticism) suggested that internal consistency might be regarded as one of the personality traits. This was consistent with the previous studies on the subject (Piedmont, McCrae, Riemann, & Angleitner, 2000; Costa & McCrae, 1997; Johnson, 2005). However, there is also a possible statistical explanation of this phenomena that is applicable to the current sample, where the frequency curves of Openness, Conscientiousness, Extraversion and Agreeableness were negatively skewed (or to the right), but the Neuroticism frequency curve was skewed positively (Table 8). Moreover, due to the skewness of the frequency curves, means in the current sample were below the median for the first four personality scores and above the median for the fifth one. The less consistent scores have more error, so they regress to the mean. Consequently, the respondents that showed less consistency scored on average lower in four former personality scores and higher in the latter, which explains the negative correlation between consistency measures and neuroticism score and positive between consistency measures and the other personality scores. This also implies that the correlation between internal consistency measures and personality scores should correlate highly with the skewness of personality score distribution. This was true for the current sample, where this correlation reached - 0.95 for JIR and - 0.85 for GPA. Furthermore, contrary to the results in the previous studies on the subject (Piedmont, McCrae, Riemann, & Angleitner, 2000; Costa & McCrae, 1997; Johnson, 2005) neither of the consistency measures loaded on any of the personality factors in the factor analysis. Additionally, the quality of the factor structure (expressed among others by the KMO

measure) rose simultaneously with the internal consistency. Above findings suggest that the relationship between internal consistency and personality scores in the current sample stems mainly from the skewness of the scores' distribution, and cannot be regarded as a proof that internal consistency is an individual difference or personality trait. This allowed to use the internal consistency measures as the indices of the protocol validity in the current study.

Table 8. Means, Modes and SD of the personality scores

| | Personality Measure | | | | |
|---|---|---|---|---|---|
| | O | C | E | A | N |
| Mean | 3.92 | 3.47 | 3.50 | 3.58 | 2.69 |
| Median | 3.95 | 3.50 | 3.55 | 3.65 | 2.65 |
| SD | .56 | .68 | .77 | .60 | .77 |
| Skewness | -0.46 | -0.22 | -0.38 | -0.48 | .23 |

The mean for Jackson's Individual Reliability Coefficient in the present sample was .73 This figure is considerably lower than .83 reported by Johnson (2005) in the similar sample which may suggest that the internal consistency in the current sample was poor. Moreover, the minimum acceptable JIR value of .30 proposed by Jackson (1977) and used by Johnson (2005) was not met by 18,843 protocols (10.3% of the sample) compared to .2% in the Johnson's (2005) sample. As JIR relies on the reliability of the instrument's scales, the lower level of JIR values in the current sample may be explained by its shorter scales in comparison to those in Johnson's (2005) 240 items questionnaire. However, it might also be a sign of the problem with scales' consistency. Before the cut-off point on the JIR measure can be recommended the consistency of the scales in the current instrument should be verified.

The average for Goldberg's antonym coefficient in the current sample (with the sign reversed) was .61. This result, higher than .47 reported in the comparable sample by Johnson (2005), might be caused by the relaxed rules used to choose the psychometric antonyms in this study. There were 6,056 protocols (3.3% of the sample) that fall below the minimum GPA value (.03) proposed by Johnson (2005). Accepting protocols with psychometric

antonyms consistency measure so close to zero may be regarded as too much tolerance for inconsistency, however the relaxed rules for selecting the psychometric antonyms in the current study made it necessary to maintain maximum wariness and use liberal norms.

Interestingly, the correlation between GPA and JIR was only 0.31, which suggests that those two measures are not interchangeable but rather complementary.

***Misrepresentation***

The examination of the KMO, the Total Variance Explained values and alpha-reliability of 5 personality scales on different levels of the Social Desirability Coefficient showed a clear and steady decrease in the quality of the factor model and reliability of the scales with an increase of SDC. This is consistent with the assumption underlying the SDC, that some participants are more concerned with the social appropriateness of their responses and maintaining their 'good image' than with the item content. However, there were no signs of decrease in the quality of the factor structure on close to zero and low levels of the SDC. It might suggest that in the current sample SDC was not a good proxy for random responding and that there were not many respondents attempting to fake bad.

Interestingly, SDC correlated highly with the personality scores. It is possible that there is a relationship between personality and readiness to misrepresent. However, this correlation might be explained in the similar way as the correlation between personality factors and internal consistency measures. SDC used in this research was, in fact, a measure of consistency – consistency with the average score on each item. Respondents that scored high on neuroticism and low on the other personality scores received low SDC and vice versa, which explains the negative correlation between SDC and neuroticism and the positive correlation between SDC and other personality scores. That is confirmed by the high correlation (-0.94) between skewness of personality score distribution and SDC/personality scores correlation coefficient.

The steadiness of the decrease in the factor model quality with the increase in SDC and lack of the comparable norms in literature did not allow to decide if SDC can be used as a proxy of protocol validity in the present instrument.

***Overlap in the exclusion rules***

There were four protocol validity indices that proved to be useful in the current sample. The total number of discarded protocols equalled 10,412 (5.7% of the sample). The number of records eliminated by each of them is showed in Table 9. Only a few protocols were invalidated by more than one exclusion rule, which proves that they those rules are highly independent and should be used simultaneously.

Table 9. Number of cases eliminated by one to three criteria

| | Cases eliminated by one or two criteria: | | | |
|---|---|---|---|---|
| | Consecutive | Patterns | Missing | Goldberg |
| Consecutive | 1465 | | | |
| Patterns | 189 | 1489 | | |
| Missing | 35 | 30 | 1562 | |
| Goldberg | 132 | 110 | 87 | 5262 |
| | Cases eliminated by three criteria: | | | |
| | Consecutive | Patterns | Missing | Goldberg |
| Consecutive and Patterns | | | 1 | 24 |
| Missing and Goldberg | 17 | 7 | | |

Note: There were 2 cases eliminated by all four criteria

## Conclusions

Invalid protocols may seriously decrease the accuracy of a personality questionnaire (Morey & Hopwood, 2004). Thus, it is essential to detect and discard them from the dataset before proceeding to the actual analysis. However, discarding invalid protocols entails the risk of accidentally removing the cases that are in fact valid, which not only senselessly decreases the sample size but may distort the research results. Four protocol validation methods that were examined in the current research proved to be useful, however criteria suggested here should be verified with the external validity indices. There are two questions that stem from this study and should be answered in the further research. Firstly, high numbers of cases that

scored very low on the Jackson Individual Reliability measure indicate a problem with scales' reliability in the current instruments, which should be verified. Secondly, the psychometric explanation of the relation between personality score distribution and internal consistency measures that were proposed here should be investigated.

## References

Archer, R. P., & Elkins, D. E. (1999). Identification of random responding on the MMPI–A. *Journal of Personality Assessment*, 73, 407–421.

Buchanan, T. (2002). Online Assessment: Desirable or Dangerous? *Professional Psychology: Research and Practice*, *33*, pp. 148-154.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, *90*, pp. 125-145.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a Five-Factor Personality Inventory for use on the Internet. *European Journal of Psychological Assessment*, *21*, pp. 115-127.

Buckley, N., & Williams, R. (2002). Testing on the web - Response patterns and image management. *Selection & Development Review*, *18*, pp. 3-8.

Caspi, A., & Gorsky, P. (2006). Online Deception: Prevalence, Motivation, and Emotion. *CyberPsychology & Behavior*, *9*, pp. 54-59.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, pp. 245–276.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-RTM) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1997). Stability and change in personality assessment: The revised NEO personality inventory in the year 2000. *Journal of Personality Assessment*, *68*, pp. 86–94.

Costa, P. T., & McCrae, R. R. (2005). The revised NEO Personality Inventory (NEO-PI-R). In S. Briggs, J. Cheek, & E. Donahue, *Handbook of Adult Personality Inventories.*

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

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that aVect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology*, *48*, pp. 82-98.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, pp. 84–96.

Gosling, S. D., Gaddis, S., & Vazire, S. (2007). Personality Impressions Based on Facebook Profiles. *Proceedings of the International Conference on Weblogs and Social Media.* Boulder, CO.

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*, pp. 93-104.

IPIP.org. (2009, 1 1). *"Validity" Indices for IPIP Measures*. Retrieved 1 17, 2009, from International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences: http://ipip.ori.org/newValidity.htm

IPIP.org. (2009, 1 1). *A Comparison between the 5 Broad Domains in Costa and McCrae's NEO Personality Inventory (NEO-PI-R) and the Corresponding Preliminary IPIP Scales Measuring Similar Constructs*. Retrieved 1 14, 2009, from International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences: http://ipip.ori.org/newNEO_DomainsTable.htm

Jackson, D. N. (1977). *Jackson Vocational Interest Survey manual.* Port Huron, MI: Research Psychologists.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*, pp. 103–129.

Johnson, J. A. (2002). Effect of construal communality on the congruence between self-report and personality impressions. *Personality Judgments: Theoretical and Applied*

*Issues*. Invited symposium for the 11th European Conference on Personality, Jena, Germany.

Knapp, F., & Heidingsfelder, M. (2001). Drop-out analysis: Effects of the survey design. *Dimensions of Internet Science*, pp. 221-230.

Kurtz, J. E., & Parrish, C. L. (2001). Semantic Response Consistency and Protocol Validity in Structured Personality Assessment: The Case of the NEO–PI–R. *JOURNAL OF PERSONALITY ASSESSMENT*, *16*, pp. 315-332.

Morey, L. C., & Hopwood, C. J. (2004). Eficiency of a strategy for detecting back random responding on the Personality Assessment Inventory. *Psychological Assessment*, *16*, pp. 197-200.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, *78*, pp. 582-593.

Reips, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. *Psychological Experiments on the Internet*, pp. 89-117.

Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO–PI–R: Development and initial validation. *Journal of Personality Assessment*, *68*, pp. 127–138.

SIGMA Assessment Systems, Inc. (2008). *Psychometric Properties*. Retrieved January 10, 2009, from JIVIS.COM: http://www.jvis.com/about/psychomet.htm

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The Developmental Psychometrics of Big Five Self-Reports: Acquiescence Factor Structure, Coherence, and Differentiation From Ages 10 to 20. *Journal of Personality and Social Psychology*, *94* (4), pp. 718-737.

Stillwell, D. (2009, 1 1). *MyPersonality Research*. Retrieved 1 1, 2009, from Mypersonality.org: http://mypersonality.org/research/interested-in-collaborating/

## Index of Tables



## Index of Figures

**Appendix 1.** Pairs of items with the highest negative correlation used in Goldberg's Psychometric Antonym Coefficient

| Correlation | Item 1 | Item 2 |
|---|---|---|
| -0,67 | 30 Dislike myself | 11 Feel comfortable with myself |
| -0,53 | 69 Find it difficult to approach others | 10 Make friends easily |
| -0,54 | 18 Avoid contact with others | 83 Feel comfortable around people |
| -0,63 | 30 Dislike myself | 97 Am very pleased with myself |
| -0,56 | 44 Tend to vote for conservative political candidates | 51 Tend to vote for liberal political candidates |
| -0,64 | 58 Leave things unfinished | 65 Finish what I start |
| -0,52 | 69 Find it difficult to approach others | 73 Am skilled in handling social situations |
| -0,54 | 69 Find it difficult to approach others | 83 83Feel comfortable around people |
| -0,54 | 69 Find it difficult to approach others | 93 Start conversations |
| -0,63 | 74 Do not like art | 91 91Believe in the importance of art |
| -0,52 | 59 Don't like to draw attention to myself | 3 Do not mind being the centre of attention |
| -0,51 | 47 Am not easily frustrated | 17 Get stressed out easily |
| -0,51 | 29 Don't talk a lot | 93 Start conversations |
| -0,51 | 69 Find it difficult to approach others | 43 Talk to a lot of different people at parties |
| -0,51 | 83 Feel comfortable around people | 79 Retreat from others |

**Appendix 2.** Detailed description of the Mypersonality Personality Questionnaire Application

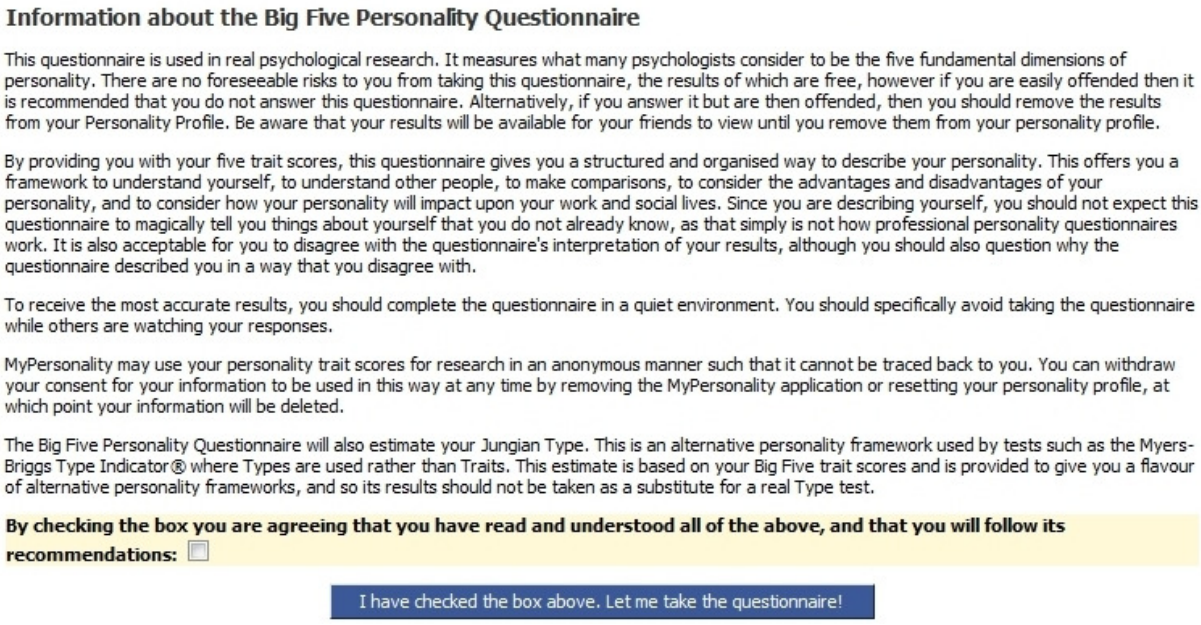
At the time of this research, Facebook users could find a link to the Mypersonality web-questionnaire in their friend's profiles or in the Facebook applications catalogue (categories: education and dating). After adding the web-questionnaire application respondents saw a webpage with a brief description of the Big Five Personality Questionnaire and Big Five Model, a short instruction regarding the optimal test-taking conditions and information that their score may be used in academic research (screen 1). Before proceeding to the questionnaire participants had to check the box next to the sentence "By checking the box you are agreeing that you have read and understood all of the above, and that you will follow its recommendations".

On the next page (screen 2) there was a short instruction on how to fill the questionnaire, together with an option to select the number of items that one was willing to answer (from 20 to 100 in increments of 10) and the default, 20 items, questionnaire. If the respondent selected different questionnaire length, the page was reloaded and the selected number of items was presented. In order to respond to a question the participant had to click on one of the five radio buttons labelled *Very Inaccurate*, *Moderately Inaccurate*, *Neither Accurate nor Inaccurate*, *Moderately Accurate*, or *Very Accurate*. Below the questionnaire the respondent had an option of keeping the results private or publishing them on their Facebook profile.

The following page consisted of respondent's raw scores together with information that was intended to help interpret them. It included a brief description of the meaning of each of the scales and the comparison of their score to the first 350,000 people to complete the full 100 items version of Mypersonality Big Five questionnaire (screen 3).

At the time of the research, Mypersonality.org application had over 285 thousands active monthly users from around the world, stored over 1.9 million personality self-ratings, over 200,000 friend-ratings and over 140,000 repeated self-ratings. Although the total number of personality scores in the database is large, the item level data has been being recorded only since recently and there are 550,000 of the records that consist responses to the test items. The database also contained detailed information form over 600,000 Facebook profiles of the respondents, including demographic data such as age and the country of residence, personal information such as favourite films, TV programs, political views, sexuality, interests, and a list of user's friends (Stillwell, 2009).

Screen 1:



**Information about the Big Five Personality Questionnaire**

This questionnaire is used in real psychological research. It measures what many psychologists consider to be the five fundamental dimensions of personality. There are no foreseeable risks to you from taking this questionnaire, the results of which are free, however if you are easily offended then it is recommended that you do not answer this questionnaire. Alternatively, if you answer it but are then offended, then you should remove the results from your Personality Profile. Be aware that your results will be available for your friends to view until you remove them from your personality profile.

By providing you with your five trait scores, this questionnaire gives you a structured and organised way to describe your personality. This offers you a framework to understand yourself, to understand other people, to make comparisons, to consider the advantages and disadvantages of your personality, and to consider how your personality will impact upon your work and social lives. Since you are describing yourself, you should not expect this questionnaire to magically tell you things about yourself that you do not already know, as that simply is not how professional personality questionnaires work. It is also acceptable for you to disagree with the questionnaire's interpretation of your results, although you should also question why the questionnaire described you in a way that you disagree with.

To receive the most accurate results, you should complete the questionnaire in a quiet environment. You should specifically avoid taking the questionnaire while others are watching your responses.

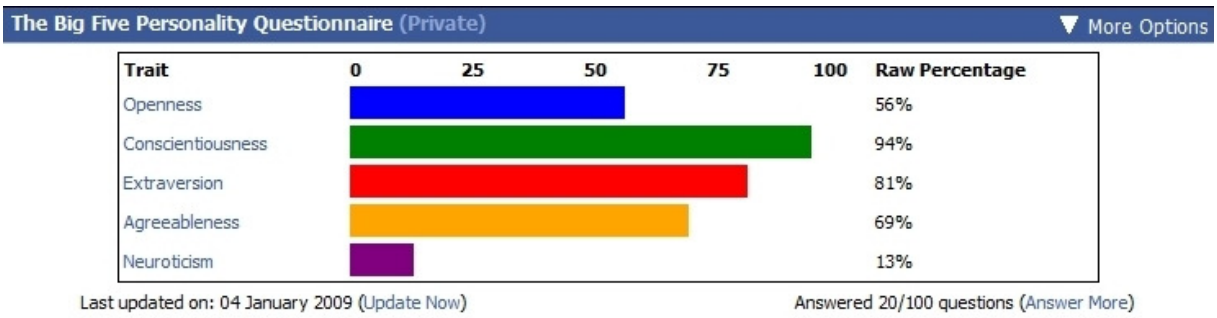
MyPersonality may use your personality trait scores for research in an anonymous manner such that it cannot be traced back to you. You can withdraw your consent for your information to be used in this way at any time by removing the MyPersonality application or resetting your personality profile, at which point your information will be deleted.

The Big Five Personality Questionnaire will also estimate your Jungian Type. This is an alternative personality framework used by tests such as the Myers-Briggs Type Indicator® where Types are used rather than Traits. This estimate is based on your Big Five trait scores and is provided to give you a flavour of alternative personality frameworks, and so its results should not be taken as a substitute for a real Type test.

**By checking the box you are agreeing that you have read and understood all of the above, and that you will follow its recommendations:** ☐

I have checked the box above. Let me take the questionnaire!

Screen 2:




## Trait Explanations

In order to interpret your raw percentages, they were compared to the first 350,000 people to complete the full MyPersonality Big Five questionnaire. This allows the way that you described yourself to be put in the context of how other people respond to the questionnaire. You should remember that there are no fundamentally good or bad personalities, as each trait description has potential advantages and disadvantages. To help you reflect on these, you have also been given some questions which ask you to consider the implications of your trait descriptions. Other people viewing your personality profile will not be able to see these.

### Openness

This trait refers to the extent to which you prefer novelty versus convention. Approximately 11.5% of respondents have a lower openness raw percentage than yours. From the way you answered the questions, you seem to describe yourself as someone who is down-to-earth and prefers things to be simple and straightforwards. You might say that it just makes life easier if things don't change unnecessarily, that the arts are of no practical use to you, and that you think tradition is more important than others do.

**Reflective question:** How do you react to change?

### Conscientiousness

This trait refers to the extent to which you prefer an organised, or a flexible, approach in life. Approximately 97.7% of respondents have a lower

Screen 3:

## The Big Five Personality Questionnaire

Below, there are phrases describing people's behaviours. Please use the rating scale to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Rate yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. If you are unsure of which response to choose (e.g. you act one way in a certain situation, and another way in a different situation), choose the response which feels most "natural" to you.

So that you can describe yourself in an honest manner, your answers to individual questions cannot be seen by others, only the overall calculation of your personality traits.

Answer 20 (2 minutes) questions (The more you answer, the more accurate your results will be. But you can always answer more later.)

| Phrase: I... | Very **Inaccurate** | Moderately Inaccurate | Neither Inaccurate nor Accurate | Moderately Accurate | Very **Accurate** |
|---|---|---|---|---|---|
| Have a vivid imagination. | ○ | ○ | ○ | ○ | ○ |
| Hold a grudge. | ○ | ○ | ○ | ○ | ○ |
| Do not mind being the centre of attention. | ○ | ○ | ○ | ○ | ○ |
| Do not like poetry. | ○ | ○ | ○ | ○ | ○ |
| Complete tasks successfully. | ○ | ○ | ○ | ○ | ○ |
| Believe that others have good intentions. | ○ | ○ | ○ | ○ | ○ |
| Avoid philosophical discussions. | ○ | ○ | ○ | ○ | ○ |
| Need a push to get started. | ○ | ○ | ○ | ○ | ○ |
| Cut others to pieces. | ○ | ○ | ○ | ○ | ○ |