

Running head: PERSON-FIT IN PERSONALITY DATA

A Person-Fit Analysis of Personality Data

Master Thesis, May 2011

Patrick Emmen

Student number: 1674854

Faculty of Psychology and Education

Department of Social and Organizational Psychology

Supervisor: Prof. dr. H. Kelderman

Second Supervisor: Dr. R. de Vries

Vrije Universiteit

Van der Boechorstraat 1

1081 BT Amsterdam

Abstract

In this paper we study the inappropriateness of response patterns on personality tests and discuss whether certain personality facets or factors can be related to responding inappropriately. We examine response patterns by using a person-fit statistic which compares the appropriateness of item score patterns with the normal response patterns within our population sample. The majority of individuals showed appropriate response patterns, and only a few revealed inappropriate response patterns. We found that some facets are more vulnerable to inappropriate responding than others. We also discovered that inappropriate responding was not specific to any particular facet, but that it could largely be described by a single factor. We found that certain facets could be related to this inappropriateness factor and the degree of inappropriateness exhibited depended on what level of the trait an individual possessed. The facets ‘altruism’, ‘anxiety’, ‘dependence’, ‘fearfulness’, ‘flexibility’, ‘forgiveness’, ‘greed avoidance’, ‘social self-esteem’ and ‘unconventionality’, as well as the factors ‘Honesty-Humility’ and ‘Extraversion’ were predictors of inappropriateness of response.

A Person-Fit Analysis of Personality Data

Master Thesis, May 2011

In personnel selection and psychological assessment, it is common to evaluate or diagnose individuals using psychological measurement instruments. Individuals used to be assessed by pen-and-paper, but nowadays computerized instruments are administered online which means more individuals can be assessed in less time and test results are often generated automatically. Measurement instruments, however, can have quite different degrees of accuracy, validity and importance to different individuals (Ghiselli, 1972). It is possible, for example, for the test score of a particular individual to be inappropriate, even if a test has satisfactory measurement properties in the general population (Dragow & Guertler, 1987). When it comes to individual assessment, it is critically important that one has confidence in the test scores used, because the decision that is based on this assessment can have important consequences for an individual. Therefore, before one interprets an individual's test score, one must first determine whether his or her response pattern is appropriate.

It is possible for the test scores of particular individuals to be similar, while their response patterns reveal a different story. Often, after a test has been administered, a single total test score is interpreted and little is done to evaluate an individual's response pattern (Rudner, 1983; Tellegen, 1988). For example, on a 20-item test, a total test score of 10 indicative responses can be obtained in 184,756 different ways (Harnisch & Linn, 1981). Different ways correspond to different response patterns, some of which do not seem logical from a substantive point of view. For example, relatively indicative answers are given when a low level of the trait is more likely. In cases such as these, an analysis of response patterns might reveal different information about the individual than an interpretation of his or her test score would suggest (Meijer & Sijtsma, 1995). Therefore, an analysis of the item score pattern is strongly recommended, particularly in situations

where individual-decision making is important (Drasgow, Levine, & McLaughlin, 1987; Parsons, 1983).

One type of research which analyzes typical and atypical item scores patterns is person-fit research (e.g., Emons, 2008; Meijer & Sijtsma, 2001). By using person-fit statistics, one is trying to detect individuals whose item score patterns on a test deviate from what is expected when compared to a test model of the normal response patterns within the population of interest. Even though person-fit statistics are sensitive to inappropriate response patterns, finding these patterns does not explain why an individual has responded in a particular way (Meijer & Sijtsma, 1995). Inappropriate response patterns can only give an indication that the respondent's response behaviour may have been influenced or threatened by factors that weren't measured by the test (Emons, Meijer, & Sijtsma, 2002). For example, personal, cognitive, and psychological factors may have affected response behaviour. At the same time, different types of inappropriate response behaviour can produce similar kinds of item response patterns (Meijer, Muijtjens, & van der Vleuten, 1996). For example, a response pattern that consists of correct answers to difficult questions and incorrect answers to easy ones could indicate, on the one hand, that an individual is cheating but, on the other, that the person might simply be more nervous at the beginning of the test.

Besides the particular trait and the ability level which is being measured by a test, a variety of other factors can also affect response behaviour in different ways (Dodeen & Darabi, 2009). Schmitt, Chan, Sacco, McFarland & Jennings (1999) examined individuals who responded inappropriately and found that inappropriateness can be related to gender, and gender differences can be associated with personality differences. For instance, females are more conscientious when completing tasks at all levels of difficulty. Levine and Rubin (1979) associated inappropriate responses with the carelessness of individuals who ignored items in reverse-word order and with individuals who had misgrid their answers

on an optical scanning sheet by putting the answer to Item 10 in the Item 11 box and then subsequently putting all the answers in the wrong boxes. Individuals who make random responses (e.g., such as those who are not motivated to take the test) and those who make exceptionally creative responses threaten the measurement of the appropriateness of response patterns. More general factors that could influence response making are: language deficiencies, a lack of education, a poorly developed sense of self, unfamiliarity with computers or an unusual instructional history (Levine & Rubin, 1979).

Inappropriate responding, however, can also depend heavily on the type of test used (Chernyshenko, Stark, Chan, K-Y., Drasgow, & Williams, 2001). In cognitive testing, for example, a respondent could have illicit foreknowledge by already possessing the test in question and memorizing some of the very difficult questions or cribbing answers from another test. In personality testing, inappropriate response patterns such as these, however, seem to be less obvious. In cognitive testing response patterns consist of correct or incorrect answers, while in personality testing response patterns often consist of answers which indicate the level of agreement to a Likert questionnaire item. In personality testing, a response pattern is inappropriate when relatively indicative answers are given when a low level of the trait is more likely and when relatively non-indicative answers are given when a high level of the trait is more likely. Because of the different nature of the type of test used, additional sources of person-misfit may arise in personality testing (Reise & Flannery, 1996).

In personality testing, response biases such as the extreme response bias, can influence the occurrence of inappropriate response patterns. Extreme response bias is type of bias by which a respondent has a tendency to use the extreme alternatives on questionnaires. Individuals who show this tendency are often identified in tests which employ items requiring the subject to respond along a Likert scale, such as 'strongly agree' to 'strongly disagree' (Hamilton, 1968). Individuals who have an extreme response

bias can produce inappropriate response patterns, because their response patterns will deviate from the expected population norm. Mittelhaeuser (2009) found, by inspecting the ten most inappropriate response patterns on an experimental scale for measuring distressed (Type D) personality, that extreme response style is indeed a possible explanation for most inappropriate response patterns. Naemi, Beal and Payne (2009) examined the type of individuals who are most likely to exhibit extreme responses and found that they often complete surveys quickly, are intolerant of ambiguity or are simplistic thinkers. Some studies suggest that extreme response bias is related to anxiety, and that high-anxiety individuals significantly choose more extreme response options than low-anxiety individuals do (Berg & Collier, 1953; Lewis & Taylor, 1955). There is, therefore, a possibility that anxiety is related to inappropriate responding because anxious people can produce extreme response patterns. Birenbaum (1986) found that respondents could falsify their responses on an anxiety test but, at the same time that their feelings of anxiety affected their cognitive functioning. This effect of anxiety on cognitive functioning contributed to the degree of inappropriate responding. Birenbaum (1986) also indicated that compulsiveness could affect test performance and cause inappropriate response patterns to be made. Compulsiveness, for example, is often characterized by a preoccupation with organizing and with perfectionism, both of which interfere with task completion. Task-irrelevant cognitions, resulting from anxiety and some elements of compulsiveness, could cause inappropriate responding (Sarason, 1984).

While some people prefer to mark extreme ends of the response scales, others have a tendency to over-report good behaviour and underreport bad. This tendency to answer in a manner which denies socially undesirable traits and claims socially desirable traits is called the social desirability bias (Nederhof, 1985). For example, responses to questions about feelings of self-worth are often inflated, while responses to questions about illegal acts are often deflated. It also seems that people often choose extreme response options to

questions which they consider particularly important and, at the same time, as questions become increasingly embarrassing, the probability of sincerity is expected to decrease (Emons, 2008; Woods, Oltmanns, & Turkheimer, 2008). The concern about evaluation - being anxious about whether one is negatively or not positively evaluated, can lead to task-irrelevant cognitions. These irrelevant cognitions can result in inappropriate responding (Sarason, 1984). In the personality assessment context, responding in a socially desirable way plays a critical role when someone's personality is being tested. Some individuals are aware of the fact that certain personality profiles represent the key to success in some particular jobs, particularly in a high-stakes testing situation such as the job selection process. Krahe (1989) found that individuals have the ability to fake specific personality profiles. Faking personality profiles and socially desirable responding could result in inappropriate responding to personality domains that are sensitive to participants' interpretations of social desirability and importance.

In this paper we examine inappropriate responding in personality testing. First, we examine the occurrence of inappropriate response patterns, as we have no existing knowledge about the percentage of inappropriate response patterns and the degree of this inappropriateness (Hendrawan, Glas, & Meijer, 2005; Molenaar & Hoijsink, 1990). Second, we examine whether facets and factors differ in the degree of inappropriateness of response. Different facets may have quite different degrees of accuracy, validity and importance to different individuals (Ghiselli, 1972). Besides the degree of inappropriateness we are also interested in the percentage of inappropriate response patterns on each facet. For example, some facets are more vulnerable to extreme and socially desirable responding than others. To examine the percentage of inappropriate response patterns we first try to make a clear distinction between appropriate and inappropriate response patterns and then we compare the percentages of inappropriate response patterns on each facet. In line with the inappropriate response patterns we

examine if there are respondents in our population sample who produce extreme response patterns. Extreme response patterns are important, because it appears that they may be able to offer an explanation for most of the inappropriate response patterns.

The type of inappropriate response behaviour can differ on each facet, because each facet measures a unidimensional construct. How does the empirical structure of inappropriateness scores correspond to each facet? We examine this by investigating the dimensionality of inappropriateness scores. Finally, we analyze whether specific facets or factors are related to inappropriate responding. Which - if any - facets or factors are related to inappropriateness? We include the factor-level scores because the intercorrelations of the factor-level scales are lower than the facet-level scales (Lee & Ashton, 2006). We suspect that the facet 'creativity' can be related to inappropriate responding as it already has been suggested to do so (Levine & Rubin, 1979). At the same time, this can also account for the facet 'anxiety', because it is related to extreme response bias which, in turn, is linked to inappropriate response behaviour. The facets 'perfectionism' and 'organization' as these traits are associated with compulsive personalities. Finally, the facets related to the factor 'Honesty-Humility' are suspected to relate to inappropriateness as these facets could be sensitive to the urge to make socially desirable responses.

Method

Sample

We used a sample taken from the study of De Vries and Van Kampen (2010). De Vries and Van Kampen (2010) approached 2000 Dutch adults by email asking them to participate in three questionnaire sessions. These sessions were spaced at intervals of two weeks and participants were paid for their participation. The 2000 Dutch adult citizens were approached by a large-scale national ISO-certified and representative internet panel

consisting of approximately 20,000 panel members (De Vries & Van Kampen, 2010, p. 247). The first questionnaire session consisted of the Revised HEXACO Personality Inventory (HEXACO-PI-R) and background variables ($N = 1,377$ responses). We have only used the data collected in this first session. Respondents with low within-person standard deviations ($<.70$) and long (>30) strings of the same answer were removed from the sample by De Vries and Van Kampen (De Vries & Van Kampen, 2010, p. 247). This resulted in 1,352 valid responses (48% women). The mean age of the participants was 47.5 years ($SD = 14.9$) and their educational levels ranged from primary education (2.1%), lower-level secondary education (17.2%), higher-level secondary education (16.5%), lower-level tertiary education (6.1%), medium-level tertiary education (24.6%), higher-level tertiary education (20.0%), to university level education (13.4%).

Measurements

The participants completed a 'self-report' version of the Dutch HEXACO Personality Inventory-Revised (HEXACO-PI-R) (De Vries, Lee, & Ashton, 2008). The Dutch HEXACO-PI-R is a translation of the English version of this self-report inventory. The HEXACO-PI-R is based on the six-dimensional HEXACO personality structure proposed by Lee & Ashton (2004, 2006). The six factors that define the HEXACO-PI-R are Honesty-Humility (H), Emotionality (E), eXtraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O). The full-length Dutch HEXACO-PI-R consists of 200 questions - all of which use a 5-point Likert response scale. Each factor consists of four facets which assess distinct traits within a factor. One interstitial facet was added: Altruism. Evidence supporting the reliability and validity of the Dutch HEXACO-PI, which is a very similar version of the Dutch HEXACO-PI-R, is reported in de Vries, Lee, & Ashton (2008).

Data analysis

Selecting a person-fit statistic. The goal of person-fit analysis is to distinguish persons whose item score patterns on a test are inappropriate, given the other normal item score patterns. A person-fit statistic in itself is designed to identify item-score vectors that may be inappropriate and to calculate the corresponding inappropriateness score. The usefulness of a person-fit statistic depends on the degree to which inappropriate item-score vectors are detected and whether the detection of this inappropriate item score vector is correct (Emons, 2008). In other words, the detection of inappropriately responding individuals must be valid and the rate of false positives should stay low.

A few studies have compared different person-fit statistics (Karabatsos, 2003; Meijer & Sijtsma, 2001). Karabatsos (2003) compared 36 person-fit statistics and found that the non-parametric $U3$ person-fit statistic belonged to the four best performing person-fit statistics. The $U3$ -statistic (Van der Flier, 1980, 1982) is a useful statistic for evaluating the appropriateness of item-score patterns. Commonly, a fixed percentage of the most atypical item-score vectors is selected for further examination (Emons et al., 2002). For example, if most of the $U3$ -values are low, selecting the 5% highest $U3$ -values to identify atypical item-score vectors is suggested (Emons, Sijtsma, & Meijer, 2005). For long tests and items with high discrimination power, the detection rate of the $U3$ -statistic for inappropriate item-vectors is satisfactory (Meijer, Molenaar, & Sijtsma, 1994). The $U3$ -statistic, however, is only applicable to dichotomous items, but our data consists of responses to a 5-point response scale. Emons (2008) proposed making a generalization of Van der Flier's (1980) $U3$ person-fit statistic which can be used for polytomous items. The $U3^p$ -statistic (Emons, 2008) is the polytomous variant of the $U3$ -statistic. The p in $U3^p$ stands for polytomous. The $U3^p$ -statistic is, therefore, suitable for the 5-point response scale items of the HEXACO-PI-R. For each respondent on every facet, an inappropriateness score is calculated, with a minimum value of 0 indicating that a

response pattern is highly appropriate, and a maximum value of 1 indicating that a response pattern is highly inappropriate. In other words, a relative high inappropriateness score indicates that the response pattern of a particular individual deviates substantially from the other response patterns found in the population of interest. The $U3^p$ -statistic is applied to each of the facets separately, because each of the 25 facets of the HEXACO-PI-R tap a unidimensional construct (Maydeu-Olivares, 2005). For example, respondent 253 has an inappropriateness score of .23 on the facet ‘perfectionism’, which indicates that his or her response pattern on this particular facet is not perfectly appropriate. The $U3^p$ -statistic cannot calculate an inappropriateness score for individuals with extreme response patterns. Extreme response patterns are response patterns on a particular facet where all the questions are answered indicatively or they are all answered non-indicatively. Details of the $U3^p$ -statistic can be found in the Appendix. For application of the $U3^p$ -statistic and for all our analyses we used the R system for statistical computing (R Development Core Team, 2009).

Results

For 1,352 respondents an inappropriateness score was calculated on every facet. 33,662 inappropriateness scores were calculated in total: no inappropriateness score could be calculated for 138 extreme response patterns. An inappropriateness score of 0 was given for extreme response patterns. The average inappropriateness scores over the 25 facets of all respondents were computed, which resulted in very low average inappropriateness scores ($M = .10$, $SD = .07$). In Table 1 the frequencies of the average inappropriateness scores of the respondents are reported. The distribution of average inappropriateness scores is positively skewed, which means that, on average, there are only a few respondents who produced relatively inappropriate patterns. Most of the respondents showed relatively appropriate response patterns. It is suggested that when most of the

inappropriateness scores are low, that a selection of approximately 5% percent of the highest inappropriateness scores is made to identify atypical item-score vectors (Emons et al., 2005). We decided to calculate an average inappropriateness score over the 5% percent highest average inappropriateness scores. This resulted in an average inappropriateness score of .30. We decided to take this value as a cut-off score to indicate whether a respondent is responding appropriately or inappropriately. Nearly all respondents (98%) had an average inappropriateness score lower than .30 and only a minority (2%) had an average inappropriateness score higher than .30. These results indicate that there are only a few respondents with a high average inappropriateness score. A high average inappropriateness score could result from consistent inappropriate responding, or by extreme inappropriate responding on some facets and appropriate responding on others.

In Table 2 the descriptive statistics of the HEXACO-PI-R and the inappropriateness scores are reported. The 25 facets that were analyzed consisted of eight items, therefore the inappropriateness scores are comparable. Respondents had, on average, higher inappropriateness scores on the facets ‘fairness’ (.16), ‘inquisitiveness’ (.16), ‘aesthetic appreciation’ (.14), ‘sincerity’ (.12), and ‘sociability’ (.12) than on other scales. On the facets ‘dependence’ (.08), ‘flexibility’ (.08), ‘forgiveness’ (.08), ‘gentleness’ (.08), and ‘patience’ (.08) respondents had on average lower inappropriateness scores than on others. The only facet that elicits almost no inappropriate responding is ‘social self-esteem’ (.01). The factor ‘Openness to Experience’ (.13) had on average higher inappropriateness scores in comparison to the other factors. Table 2 indicates that different facets and factors suffer to different degrees from inappropriate responding.

The number of inappropriate response patterns and the number of extreme response patterns on each facet are reported in Table 3. We use the previously calculated cut-off score ($\geq .30$) to indicate whether a respondent’s response pattern is inappropriate. The total number of response patterns is 33,800 and 2,113 response patterns had an

inappropriateness score of .30 or higher. The percentage of inappropriate response patterns on each facet is derived from the total number of response patterns. The facets ‘fairness’ (0.6%), ‘inquisitiveness’ (0.6%), and ‘aesthetic appreciation’ (0.5%) showed relative high percentages of inappropriate response patterns, while the facet ‘social self-esteem’ (0.0%) showed a very low percentage of inappropriate response patterns. This finding suggest that some facets are more vulnerable than others to inappropriateness of response.

In total 138 extreme response patterns were found. The facets ‘fairness’ (28), ‘modesty’ (20), ‘greed avoidance’ (15), ‘altruism’ (14), ‘social self-esteem’ (12), and ‘organization’ (12) contained a relatively high number of extreme response patterns. Almost all extreme response patterns consisted of non-indicative answers, which means that a respondent scores low on the trait that is being measured. For the facet ‘forgiveness’ (8), the opposite was true. The extreme response patterns found on this facet consisted of indicative answers, which means that a respondent scores high on this trait. These findings suggests that there are individuals that show extreme response bias and that depending on the trait that is being measured, individuals either respond extremely indicatively or extremely non-indicatively.

Before we began with the principal component analysis, we normalized the inappropriateness scores of the 25 facets using a rank-transformation (Conover & Iman, 1981). Although transformed, the facets ‘altruism’, ‘dependence’, ‘diligence’, ‘fairness’, ‘forgiveness’, ‘greed avoidance’, ‘modesty’, ‘organization’, ‘sentimentality’, ‘sincerity’, ‘social boldness’ and ‘social self-esteem’ did not become perfectly normally distributed. After the rank-transformation, a principal component analysis (PCA) was used to analyse the empirical structure of inappropriateness scores and to see whether the structure differed on each facet. The scree test revealed a clear break after the first component. The first ten eigenvalues are 7.96, 1.02, 0.94, 0.89, 0.86, 0.84, 0.81, 0.79, 0.79, and 0.75. The

solution was obliquely rotated using a promax rotation procedure. The rotated solution yielded one interpretable common component with a majority of substantial loadings ($>.30$) except for the facet ‘social self-esteem’, see Table 4. This rotated component accounted for 32% of the inappropriateness scores variance. The facet ‘social self-esteem’ had a lower loading ($<.30$) than other facets, which was the only indicator for a second component. We have seen from Table 2 that the facet ‘social self-esteem’ shows very little inappropriate responding. The second factor seems to be a difficulty factor (McDonald, 1985, p. 198). Thus, the principal component analysis showed a unidimensional structure of the inappropriateness scores of almost all the facets. The only exception was found on the facet ‘social self-esteem’, which only related to the inappropriateness scores of other facets to a very small extent.

A multiple regression was conducted to examine whether specific facets can be related to inappropriate responding. The predictors were 25 facets containing the HEXACO-PI-R scores of all the respondents, while the criterion variable was the component scores of the first factor computed from rotated principal component analysis. The linear combination of these facets was significantly related to inappropriateness, $F(25, 1326) = 6.86, p < .001$. The multiple R-squared was .12, indicating that approximately 12% of the variance of the inappropriateness component in the sample could be accounted for by the linear combination of personality scores. In Table 5, it is clear that the facets ‘fearfulness’, ‘forgiveness’ and ‘flexibility’ which show a positive magnitude, and the facets ‘altruism’, ‘anxiety’, ‘greed avoidance’, ‘social self-esteem’ and ‘unconventionality’ which show a negative magnitude, are predictors of inappropriateness of response. The highest magnitudes are found on the facets ‘altruism’, ‘anxiety’, ‘social self-esteem’, and ‘unconventionality’. Judgments about the relative importance and the direction (i.e., positive or negative) of these predictors are difficult because they are correlated, however, the correlation coefficients give support to some extent. We also regressed the factor-level

scores on the component scores, see Table 6. The results in Table 6 show that the factors ‘Honesty-Humility’ and ‘Extraversion’ which show a negative magnitude can be related to inappropriateness of response. The factor ‘Agreeableness’ only shows a significant regression coefficient with inappropriateness of response, while the factor ‘Conscientiousness’ only shows a significant correlation coefficient. The factors ‘Emotionality’ and ‘Openness to Experience’ are not related to inappropriate responding. In these analyses the inappropriateness scores over the full range were analyzed and the results suggest that some facets and factors can be related to inappropriateness of response.

In the linear regression analysis, the inappropriateness scores over the full range were analyzed. To study the influence of *extremely* inappropriate responding, we employed our cut-off score of .30. We divided the sample between appropriate responding respondents and inappropriate responding respondents. We used this new binary variable to conduct a logistic regression analysis to see if the same, or other, facets or factors can be related to extreme inappropriate responding. The results of the logistic regression are reported in Table 7. The facets ‘altruism’ and ‘dependence’ are predictors of extreme inappropriateness of response. Other facets only show a significant regression coefficient: ‘creativity’, ‘fairness’, ‘fearfulness’, ‘greed avoidance’ and ‘flexibility’. The factor-level scores were also regressed on the binary variable, see Table 8. The factor ‘Agreeableness’ can be related to extreme inappropriateness of response, however, the odd ratio of this factor is very close to 1. The results of the logistic regressions suggest, when a clear distinction is made between appropriate and inappropriate responding, the facets ‘altruism’ and ‘dependence’ and to a lesser extent the factor ‘Honesty-Humility’ are predictors of extreme inappropriateness of response.

Discussion

A person-fit analysis in personality data indicated that only a small percentage of the respondents showed inappropriate response patterns. Besides inappropriate response patterns, we also found extreme response patterns. Additionally, we found that different facets and factors suffer to different degrees from inappropriate responding. The degree of inappropriateness exhibited depended on the level of the trait an individual possessed. Furthermore, inappropriate responding was not specific to a particular facet and conformed to a unidimensional construct. Eight facets and two factors were related to inappropriateness of response, and two facets and one factor were related to extreme inappropriate responding.

A relative small percentage of the respondents showed inappropriate response patterns, but this does not mean that these inappropriate responses patterns are wrong: they only give an indication that a test score might not be representative of that particular individual on that particular facet. In diagnostic settings extra attention is recommended when a response pattern is inappropriate, because these patterns are highly unlikely and can reveal important information about the response behaviour of that particular individual.

In our sample, we also found individuals who produced extreme response patterns. The nature of extreme response patterns differed from indicative to non-indicative, depending on the trait being measured. Extremes response patterns are highly unlikely, and, based on a previous study, we assumed that extreme response patterns are inappropriate (Mittelhaeuser, 2009). Extreme response patterns were mostly found on the factor 'Honesty-Humility', and these patterns usually consisted of non-indicative answers. Non-indicative answers mean that an individual possesses a very low level of the trait. Our finding that inappropriate responding might be related to individuals who possess a low level of a particular trait is similar to what Mittelhaeuser (2009) found. The facet

‘fairness’ had the highest number of individuals who produced non-indicative extreme response patterns, followed by the facets ‘modesty’ and ‘greed avoidance’. Low scorers on ‘fairness’ have a tendency to be involved in fraud and corruption, while low scorers on ‘modesty’ consider themselves entitled to privileges that others do not have. On the facet ‘greed avoidance’ a low score means that an individual wants to display wealth and privilege. The type of individuals who showed extreme response patterns on the factor ‘Honesty-Humility’ can be characterized as individuals who tend to feel a strong sense of self-importance and who are tempted to ‘bend’ laws for personal profit. These individuals might exaggerate their response to a specific question about wealth and self-importance to emphasize their social status and wealth. This deviant response behaviour can result in inappropriate response patterns. The facet ‘forgiveness’ showed extreme response patterns that consisted of indicative answers, which means that a respondent possesses a high level of the trait being measured. High scorers on the facet ‘forgiveness’ tend to re-establish friendly relationships after having been treated badly. People who are willing to forgive are probably perceived by society as positive, therefore, high scorers might aim to get a positive evaluation. As a consequence of aiming for a positive evaluation, their response patterns might deviate from the other normal response in the population. It is possible, however, that extremely inappropriate responding individuals are individuals who lack the underlying trait structure measured by the facet (Tellegen, 1988). When individuals lack the underlying trait structure of a facet, an interpretation of a personality score is out of the question and the individual can only be flagged up as inappropriately responding.

When we focused on specific facets, we found differences in the degree and amount of inappropriate response patterns. The differences between the mean inappropriateness scores on facets may indicate that some facets attach more importance to individuals than others. It is possible, for example, that the facets of the factor ‘Honesty-Humility’ are perceived as being more important by society (i.e., evaluation apprehension) than the

facets of other factors. Besides the most extreme response patterns, the facet 'fairness' also showed a relatively high percentage of inappropriate response patterns. This suggests that the facet 'fairness' is subject to extreme socially desirable responding. The facets 'aesthetic appreciation' and 'inquisitiveness' also showed a relatively high percentage of inappropriate response patterns, which might be associated with certain aspects of the facet, for example, the number of reverse-worded items or the content of the questions. The facet 'social self-esteem' was an exception, because this facet contained extremely low inappropriateness scores. An explanation for these extremely low scores could be that most response patterns were identical and, as a result, they all looked appropriate.

We also analyzed the dimensionality of inappropriateness scores. The results indicated that there is one inappropriateness component, and that inappropriate responding could be a unidimensional construct. This suggests that the nature of inappropriateness of response was not related to specific properties of the facets. We realize, however, that different types of inappropriate response behaviour result in the same inappropriateness score. This could be the reason why we did not find facet-specific inappropriateness. An exception was found, the facet 'social self-esteem' loaded on a different component to all the other facets, but this was probably the result of the extremely low values of the inappropriateness scores.

To examine which facets or factors are related to inappropriate and extreme inappropriate responding four regression analyses were conducted. The multiple regression analyses indicated that eight facets and two factors could be related to inappropriate responding. The negative correlation corresponded with the negative regression slope found on the facets 'altruism', 'anxiety', 'greed avoidance', 'social self-esteem' and 'unconventionality', as well as on the factors 'Honesty-Humility' and 'Extraversion'. This means that the prediction of inappropriateness will decrease when a personality score increases, a finding which suggests that inappropriateness can be linked to individuals who

possess a low level of particular traits. More interestingly, this is the second time that we can relate inappropriate responding to individuals who possess a low level of these traits. The opposite is true for the facets ‘fearfulness’, ‘flexibility’, ‘forgiveness’ which all show a positive correlation with a positive regression slope. The prediction of inappropriateness on these facets will increase if an individual has a high score.

With regard to the scale description of the facets and factors: we try to give a reasonable interpretation as to how the related facets and factors might have a meaning in contributing to responding inappropriately. Individuals with low scores on the facet ‘altruism’ are not upset by the prospect of hurting others and may be seen as hard-hearted. These individuals might imagine irrelevant feelings against others, and the focus on the questionnaire could suffer which could result in inappropriate response behaviour. The ‘anxiety’ facet assesses a tendency to worry in a variety of contexts and low scorers feel little stress in response to difficulties. It is possible that these low scorers are also respondents who exhibit careless response behaviour. On the facet ‘social self-esteem’, low scorers tend to have a sense of personal worthlessness and to see themselves as unpopular. These low scorers might have a less developed sense of self and inflate their answers on questions about feelings of self-worth. The ‘unconventionality’ facet assesses one’s tendency to accept the unusual and low scorers avoid eccentric behaviour. It could be that conventional behaviour is viewed more favourably by others, so the individual decides to give socially desirable answers. Low scorers on the factor ‘Extraversion’ tend to be rather reserved, they feel awkward when they are the center of social attention and they consider themselves unpopular. As questions become increasingly embarrassing for less extravert people, it may be possible that the probability of responding sincerely might decrease which can result in inappropriate response patterns.

The facet ‘fearfulness’ assesses the tendency to experience fear. High scorers on the ‘fearfulness’ facet are strongly inclined to avoid physical harm, therefore, they might

experience strong evaluation apprehension which results in task-irrelevant response behaviour. The ‘flexibility’ facet assesses high scorers as individuals who avoid arguments and accommodate others’ suggestions, even when these may be unreasonable. High scorers on this facet may be more inclined to opt for the answer ‘agree’ than low scorers, therefore, their response pattern might not be the same as the other normal item score patterns. High scorers on the facet ‘forgiveness’ are linked to inappropriateness of response for the second time (i.e., indicative extreme response patterns and positive coefficient). Thus, in contrast with earlier results, we found that individuals who possess a high level of certain traits can also be associated with inappropriate responding.

We did two logistic regression analyses to examine the probability of extreme inappropriate responding on facet- and factor-level. Instead of eight facets, the logistic regression illustrated that two facets and one factor were related to extreme inappropriateness: ‘altruism’, ‘dependence’ and ‘Honesty-Humility’. However, the relatedness of the factor ‘Honesty-Humility’ to extremely inappropriate responding is very weak. All related facets found in the linear regression no longer showed a significant relationship to inappropriateness. The odds ratios of the two facets were different: the facet ‘dependence’ showed an odds ratio higher than 1, and the facet ‘altruism’ an odds ratio lower than 1. This means that inappropriate responding is more likely in respondents who score more highly on the facet ‘dependence’, while inappropriate responding is less likely for respondents who score higher on the facet ‘altruism’. The low scorers on ‘altruism’ are again linked to inappropriate responding. Furthermore, high scorers on the facet ‘dependence’ often seem to respond more inappropriately than low scorers. This facet assesses the need for emotional support from others and high scorers might worry about their evaluation of others. This evaluation apprehension could lead to a degree of false self-representation that could stimulate inappropriate responding.

Finally, a comparison of the predictors found in the regression analyses showed that

several facets and factors can be related to inappropriateness of response. Surprisingly, no combination of significant correlation coefficient and significant regression coefficient were found for the facets 'creativity' 'perfectionism' and 'organization'. The facet 'anxiety' and the factor 'Honesty-Humility' were - as we suggested - indeed related to inappropriateness of response. However, facets other than those suggested were related to inappropriateness. For example, the facet 'altruism' was found related to inappropriateness in both of the regression analyses. Other personality questionnaires should be analyzed to replicate our findings and to reach consensus about which facets or factors are related to inappropriate responding. The difference in the amount of significant predictors found between the two regression analyses can probably be explained by the difference in power. The logistic regression analyses, in which the two groups were extremely unequal in size, probably had less power than the multiple regression analyses. We decided, however, to do the analysis because the probability of the occurrence of extreme inappropriateness is interesting and provides complementary information. The results of the linear regression appear to be more statistically sound than the logistic regression, because we had 1,352 respondents in this analysis. However, the results of the linear regression showed multicollinearity because some predictor variables were highly correlated.

The first limitation of this paper is that the generalized $U3^p$ person-fit statistic doesn't provides us with a generally satisfactory cut-off value which indicates that a response pattern is appropriate or inappropriate. The cut-off value used in this paper was .07 higher than the cut-off value of Emons (2005), which was based on items of an intelligence questionnaire. We also assumed that inappropriate scores above .30 were inappropriate, however, we are not sure whether these response patterns are inappropriate. The decision to take this cut-off value can be explained by the fact that detecting an individual as inappropriate does not have serious consequences in the first instance for the individual being tested. It only gives an indication that this individual

might show different response behaviour than any of the other individuals. For example, one could decide to retest this individual under different circumstances (Emons et al., 2005). A second limitation concerns the detection rates of inappropriate patterns. The detection rates could differ between short and long tests (e.g., 20 items) as the detection rate of inappropriate response patterns seems to be lower on short tests than long (Reise & Allan, 1991). On the other hand, Meijer, Molenaar & Sijtsma (1994) illustrated that the detection rates of inappropriate response patterns are approximately the same for short tests as they are for long ones. A third limitation is that some inappropriateness scores on facets, even after a rank-transformation, were not normally distributed. We did not remove outliers, because there only was a small percentage of them. The main reason why these facets did not become normally distributed was because of insufficient data discrimination (i.e., many values close to zero). A fourth limitation is that the interpretations given to the facets and factors that can be related to inappropriateness of response may be questionable due to multicollinearity of the coefficients found in the regression analyses. The bivariate correlations, however, partially support the regression coefficients. Finally, we did not examine more general factors such as language deficiencies, a lack of education, a less developed sense of self, unfamiliarity with computers or unusual instructional histories that could exert influence on inappropriate responding. Although different interpretations are given to explain inappropriate response patterns, we still have to keep these general factors in mind.

A person-fit analysis is a useful tool for indicating invalid measurement and for identifying a proportion of respondents who may need additional or a different type of testing. At the same time, one must keep in mind that not every individual can be meaningfully scaled by using a personality trait (Reise & Waller, 1993; Tellegen, 1988). Assessing individuals who have responded inappropriately may lead to an inaccurate assessment their of abilities. Therefore, before interpreting the test results, we recommend

that a person-fit analysis should be carried out first to examine the fit or match between an individual's abilities and certain personality traits.

References

- Berg, I. A., & Collier, J. S. (1953). Personality and group differences in extreme response sets. *Educational Psychological Measurement, 13*, 164–169.
- Birenbaum, M. (1986). Effect of dissimulation motivation and anxiety on response pattern appropriateness measures. *Applied Psychological Measurement, 10*, 167–174.
- Chernyshenko, O. S., Stark, S., Chan, K-Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*, 523–562.
- Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*, 124–129.
- De Vries, R. E., Lee, K., & Ashton, M. C. (2008). The Dutch HEXACO personality inventory: Psychometric properties, self-other agreement, and relations with psychopathy among low and high acquaintanceship dyads. *Journal of Personality Assessment, 90*, 142–151.
- De Vries, R. E., & Van Kampen, D. (2010). The HEXACO and 5DPT models of personality: A comparison and their relationships with psychopathy, egoism, pretentiousness, immorality and machiavallianism. *Journal of Personality Disorders, 24*, 244–257.
- Dodeen, H., & Darabi, M. (2009). Person-fit: Relationship with four personality tests in mathematics. *Research Papers in Education, 24*, 115–126.
- Drasgow, F., & Guertler, E. (1987). A decision-theoretic approach to the use of appropriateness measurement for detecting invalid test and scale scores. *Journal of Applied Psychology, 72*, 10–18.
- Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement, 11*, 59–79.

- Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement, 32*, 224–247.
- Emons, W. H. M., Meijer, R. R., & Sijtsma, K. (2002). Comparing simulated and theoretical sampling distributions of the U3 person-fit statistic. *Applied Psychological Measurement, 26*, 88–108.
- Emons, W. H. M., Sijtsma, K., & Meijer, R. R. (2005). Global, local, and graphical person-fit analysis using person-response functions. *Psychological Methods, 10*, 101–119.
- Ghiselli, E. E. (1972). Comment on the use of moderator variables. *Journal of Applied Psychology, 56*, 270.
- Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin, 69*, 192–203.
- Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Education Measurement, 18*, 133–146.
- Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous mokken IRT model. *Applied Psychological Measurement, 1*, 337–352.
- Hendrawan, I., Glas, C. A. W., & Meijer, R. R. (2005). The effect of person misfit on classification decisions. *Applied Psychological Measurement, 29*, 26–44.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education, 16*, 277–298.
- Krahé, B. (1989). Faking personality profiles on a standard personality inventory. *Personality and Individual Differences, 10*, 437–443.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*, 329–358.

- Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO personality inventory: Two new facet scales and an observer report form. *Psychological Assessment, 18*, 182–191.
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics, 4*, 269–290.
- Lewis, N. A., & Taylor, J. A. (1955). Anxiety and extreme response preferences. *Educational and Psychological Measurement, 15*, 111–116.
- Maydeu-Olivares, A. (2005). Further empirical results on parametric versus non-parametric IRT modeling of Likert-type personality data. *Multivariate Behavioral Research, 40*, 261–279.
- McDonald, R. P. (1985). *Factor analysis and related models*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Meijer, R. R., Molenaar, I. W., & Sijtsma, K. (1994). Influence of test and person characteristics on nonparametric appropriateness measurement. *Applied Psychological Measurement, 18*, 111–120.
- Meijer, R. R., Muijtjens, A. M. M., & van der Vleuten, C. P. M. (1996). Nonparametric person-fit research: Some theoretical issues and an empirical example. *Applied Measurement in Education, 9*, 77–89.
- Meijer, R. R., & Sijtsma, K. (1995). Detection of aberrant item score patterns: A review of recent developments. *Applied Measurement in Education, 8*, 261–272.
- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*, 107–135.
- Mittelhaeuser, M. (2009). *Using person-fit analysis in scale revision: Application to the experimental DS scale*. Unpublished manuscript, Tilburg University.
- Molenaar, I. W., & Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika, 55*, 75–106.

- Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of Personality, 77*, 261-286.
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*, 263-280.
- Parsons, C. K. (1983). The identification of people for whom job descriptive index scores are inappropriate. *Organizational Behavior and Human Performance, 31*, 365-393.
- R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reise, S. P., & Allan, M. D. (1991). The influence of test characteristics on the detection of aberrant response patterns. *Applied Psychological Measurement, 15*, 217-226.
- Reise, S. P., & Flannery, W. P. (1996). Assessing person-fit on measures of typical performance. *Applied Measurement in Education, 9*, 9-26.
- Reise, S. P., & Waller, N. G. (1993). Traitendness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology, 65*, 143-151.
- Rudner, L. W. (1983). Individual assessment accuracy. *Journal of Educational Measurement, 20*, 207-219.
- Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46*, 929-938.
- Schmitt, N., Chan, D., Sacco, J. M., McFarland, L. A., & Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. *Applied Psychological Measurement, 23*, 41-53.
- Tellegen, A. (1988). The consistency in personality assessment. *Journal of Personality, 56*, 621-663.
- Van der Flier, H. (1980). *Vergelijkbaarheid van individuele testprestaties [Comparability of individual test performance]*. Lisse, the Netherlands: Swets & Zeitlinger.
- Van der Flier, H. (1982). Deviant response patterns and comparability of test scores.

Journal of Cross-Cultural Psychology, 13, 267–298.

Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2008). Detection of aberrant responding on a personality scale in a military sample: An application of evaluating person fit with two-level logistic regression. *Psychological Assessment, 20*, 159–168.

Appendix

Generalized $U3$ person-fit statistic

The generalized $U3$ person-fit statistic is defined in the context of Mokken scaling for polytomous item scores (e.g., Hemker, Sijtsma & Molenaar, 1995). In case of the generalized $U3$ statistic, “Let J be the number of items, with $M+1$ ordered response categories; let X_j be the random variable for the score on item j ($j = 1, \dots, J$), with possible realizations $x_j = 0, \dots, M$; and let $X_+ = \sum_{j=1}^J X_j$ be the sum score.” (Emons, 2008, p. 225). In the generalized $U3$ statistic a polytomous item is considered to be a sequence of dichotomous steps. Items with five response categories will be transformed in four dichotomous item-steps (i.e., $M = 4$). For example, when a respondent answers 4 on a 5-point response scale, the item-step score vector will be (1,1,1,0).

The other part of the generalized $U3$ statistic is denoted as follows: “Let Y denote the item-step score variable, with realizations 1 if the item step is passed, and 0 if not; let the random vector \mathbf{Y} denote the joint vector of JM item-step scores in ascending item-step difficulty. Let π_{jx_j} be the item-step difficulty, which is the population proportion of respondents with a score x_j or higher on item j , and let $\hat{\pi}_{jx_j}$ be its sample estimate. Let JM item steps be ordered and numbered by increasing difficulty $\hat{\pi}_1 \geq \hat{\pi}_2 \geq \dots \geq \hat{\pi}_k \geq \dots \geq \hat{\pi}_{JM}$, with ($k = 1, \dots, JM$); and let $\mathbf{Y} = (Y_1, \dots, Y_k, \dots, Y_{jm})$ be the corresponding ordered vector of the JM item-step scores and $\mathbf{y} = (y_1, \dots, y_k, \dots, y_{JM})$ its realization.” (Emons, 2008, p. 227).

Generalized $U3$ Person-Fit Statistic. We use the generalization of the $U3$ person-fit statistic (Van der Flier, 1980) for polytomous items as it is proposed by Emons (2008).

“For an observed vector \mathbf{y} , let

$$W(\mathbf{y}) = \sum_{k=1}^{JM} y_k \log \left(\frac{\hat{\pi}_k}{1 - \hat{\pi}_k} \right), \quad (1)$$

which is the sum of the log odds of the item-step difficulties of the steps that were passed. The polytomous generalization of $U3$, denoted by $U3^p$, is obtained by norming $W(\mathbf{y})$ as follows:

$$U3^p = \frac{\max(W|X_+) - W(\mathbf{y})}{\max(W|X_+) - \min(W|X_+)}, \quad (2)$$

With a minimum value of $U3^p$ equal to 0 indicating no misfit, and a value of 1 indicating extreme misfit. The maximum $\max(W|X_+)$ in equation 2 is obtained if and only if the X_+ easiest item steps are passed; that is,

$$\max(W|X_+) = \sum_{k=1}^{X_+} \text{logit}(\hat{\pi}_k), \quad (3)$$

Because of structural dependencies between the item-step scores, the minimum value, $\min(W|X_+)$, cannot be expressed in closed form. Therefore, $\min(W|X_+)$ was computed using a recursion algorithm.” (Emons, 2008, p. 228). Details of the recursion algorithm can be found in the appendix in Emons (2008).

Acknowledgements

The author would like to thank Prof. dr. H. Kelderman for his comments and suggestions, and Dr. R. de Vries for providing the data sample. A special thanks goes to Wilco H.M. Emons for providing the person-fit statistic. Author's address: Eerste van der Helststraat 19 II, 1073 AB Amsterdam, the Netherlands. E-mail: patrickemmen@gmail.com.

Table 1

Binned empirical distribution of average inappropriateness score of respondents over the 25 facets of the HEXACO-PI-R.

Average inappropriateness score	Frequency	%
0.0 - 0.1	808	59.8
0.1 - 0.2	438	32.4
0.2 - 0.3	78	5.8
0.3 - 0.4	24	1.8
0.4 - 0.5	3	0.2
0.5 - 0.6	0	0.0
0.6 - 1.0	1	0.1
Total	1352	100.0

Table 2

Descriptive statistics of the HEXACO-PI-R scores and inappropriateness scores.

	α	HEXACO-PI-R score		Inappropriateness score	
		M	SD	M	SD
Honesty-Humility (H)	.91	2.31	.49	.12	.10
Sincerity	.75	2.40	.57	.12	.14
Fairness	.81	2.21	.67	.16	.17
Greed Avoidance	.83	2.42	.67	.11	.11
Modesty	.80	2.20	.56	.11	.12
Emotionality (E)	.88	2.87	.44	.10	.08
Fearfulness	.71	3.14	.56	.11	.10
Anxiety	.78	2.89	.64	.11	.10
Dependence	.80	2.86	.60	.08	.09
Sentimentality	.79	2.61	.61	.11	.11
Extraversion (X)	.90	2.63	.47	.09	.07
Social Self-Esteem	.81	2.15	.52	.01	.05
Social Boldness	.84	2.88	.69	.11	.12
Sociability	.74	2.90	.61	.12	.11
Liveliness	.82	2.59	.62	.10	.11
Agreeableness (A)	.88	2.94	.41	.08	.07
Forgiveness	.85	3.27	.64	.08	.10
Gentleness	.73	2.83	.51	.08	.09
Flexibility	.62	2.96	.47	.08	.08
Patience	.74	2.73	.55	.08	.08
Conscientiousness (C)	.85	2.58	.39	.10	.08
Organization	.83	2.54	.68	.11	.13
Diligence	.74	2.58	.53	.10	.10
Perfectionism	.72	2.45	.54	.11	.11
Prudence	.71	2.76	.51	.09	.09
Openness to Experience (O)	.87	2.81	.47	.13	.09
Aesthetic appreciation	.78	2.75	.69	.14	.13
Inquisitiveness	.78	2.61	.68	.16	.15
Creativity	.74	2.89	.62	.11	.10
Unconventionality	.69	2.98	.51	.09	.09
Altruism	.75	2.11	.49	.10	.12

Note. $N = 1352$.

Table 3

Percentage of inappropriate response patterns and number of extreme response patterns on each facet of the HEXACO-PI-R.

	Inappropriate response patterns	Extreme response patterns
Honesty-Humility (H)	1.5%	69
Sincerity	0.3%	6
Fairness	0.6%	28
Greed Avoidance	0.2%	15
Modesty	0.3%	20
Emotionality (E)	1.0%	5
Fearfulness	0.3%	0
Anxiety	0.2%	2
Dependence	0.2%	1
Sentimentality	0.3%	2
Extraversion (X)	0.8%	21
Social Self-Esteem	0.0%	12
Social Boldness	0.3%	1
Sociability	0.3%	1
Liveliness	0.2%	7
Agreeableness (A)	0.5%	10
Forgiveness	0.2%	8
Gentleness	0.1%	0
Flexibility	0.1%	1
Patience	0.1%	1
Conscientiousness (C)	0.9%	15
Organization	0.3%	12
Diligence	0.2%	0
Perfectionism	0.2%	2
Prudence	0.1%	1
Openness to Experience (O)	1.4%	4
Aesthetic appreciation	0.5%	2
Inquisitiveness	0.6%	2
Creativity	0.2%	0
Unconventionality	0.1%	0
Altruism	0.3%	14

Note. $N = 1352$. Total number of response patterns is 33800.

Table 4

Promax rotated principal component loadings of the inappropriateness scores on each facet of the HEXACO-PI-R.

	Component
Honesty-Humility (H)	
Sincerity	.56
Fairness	.48
Greed Avoidance	.53
Modesty	.61
Emotionality (E)	
Fearfulness	.61
Anxiety	.59
Dependence	.55
Sentimentality	.59
Extraversion (X)	
Social Self-Esteem	.26
Social Boldness	.59
Sociability	.63
Liveliness	.59
Agreeableness (A)	
Forgiveness	.49
Gentleness	.57
Flexibility	.57
Patience	.61
Conscientiousness (C)	
Organization	.59
Diligence	.59
Perfectionism	.58
Prudence	.60
Openness to Experience (O)	
Aesthetic appreciation	.56
Inquisitiveness	.60
Creativity	.61
Unconventionality	.58
Altruism	.53

Note. $N = 1352$. Loadings $>.30$ are printed in bold.

Table 5

Summary of multiple regression analysis of the inappropriateness component on facet-level scores of the HEXACO-PI-R.

	Bivariate correlation	<i>B</i>	<i>SE(B)</i>	β
Honesty-Humility (H)				
Sincerity	-.07**	.06	.06	.04
Fairness	-.06*	.03	.05	.02
Greed Avoidance	-.11**	-.14*	.05	-.09*
Modesty	-.10**	-.03	.07	-.02
Emotionality (E)				
Fearfulness	.06*	.12*	.06	.07*
Anxiety	-.06*	-.21**	.06	-.14**
Dependence	.01	.10	.06	.06
Sentimentality	-.08**	.03	.06	.02
Extraversion (X)				
Social Self-Esteem	-.14**	-.27**	.07	-.14**
Social Boldness	-.06*	-.08	.05	-.05
Sociability	-.04	-.03	.06	-.02
Liveliness	-.07*	-.02	.06	-.01
Agreeableness (A)				
Forgiveness	.07**	.18**	.05	.11**
Gentleness	-.04	-.04	.07	-.02
Flexibility	.06*	.20**	.07	.09**
Patience	-.05	-.14*	.07	-.08*
Conscientiousness (C)				
Organization	-.07**	-.08	.05	-.05
Diligence	-.06*	.07	.07	.04
Perfectionism	-.11**	-.10	.06	-.06
Prudence	-.02	.04	.06	.02
Openness to Experience (O)				
Aesthetic appreciation	-.00	.11*	.05	.07*
Inquisitiveness	-.00	.02	.05	.01
Creativity	-.00	.15**	.06	.10**
Unconventionality	-.10**	-.31**	.07	-.16**
Altruism	-.18**	-.35**	.08	-.17**
Constant		1.40**	.48	

Note. $N = 1352$. $R^2 = .12$.

* $p < .05$. ** $p < .01$.

Table 6

Summary of multiple regression analysis of the inappropriateness component on factor-level scores of the HEXACO-PI-R.

	Bivariate correlation	B	$SE(B)$	β
Honesty-Humility (H)	-.11**	-.24**	.06	-.12**
Emotionality (E)	-.03	-.08	.06	-.04
Extraversion (X)	-.09**	-.21**	.06	-.10**
Agreeableness (A)	.02	.17*	.07	.07*
Conscientiousness (C)	-.10**	-.16*	.07	-.06
Openness to Experience (O)	-.03	-.03	.07	-.01
Constant		1.35**	.39	

Note. $N = 1352$. $R^2 = .03$.

* $p < .05$. ** $p < .01$.

Table 7

Summary of logistic regression analysis for inappropriate respondents ($n = 68$) and appropriate respondents ($n = 1284$) on facet-level scores of the HEXACO-PI-R.

	Bivariate correlation	B	$SE(B)$	e^B
Honesty-Humility (H)				
Sincerity	-.00	.12	.42	1.12
Fairness	.04	1.12**	.36	3.07**
Greed Avoidance	-.07	-1.01*	.41	.37*
Modesty	-.02	.52	.50	1.69
Emotionality (E)				
Fearfulness	.09	.96*	.44	2.62*
Anxiety	.02	-.49	.42	.61
Dependence	.06*	.82*	.41	2.27*
Sentimentality	-.03	-.32	.46	.72
Extraversion (X)				
Social Self-Esteem	-.06*	-.45	.57	.64
Social Boldness	-.04	-.59	.43	.56
Sociability	-.04	-.46	.44	.63
Liveliness	-.05	.08	.47	1.09
Agreeableness (A)				
Forgiveness	-.07**	-.49	.36	.61
Gentleness	-.08**	-.96	.52	.38
Flexibility	-.00	1.08*	.51	2.94*
Patience	-.05	-.18	.51	.84
Conscientiousness (C)				
Organization	-.02	-.17	.34	.84
Diligence	-.01	.16	.52	1.17
Perfectionism	-.00	.29	.42	1.33
Prudence	-.05	.03	.46	1.03
Openness to Experience (O)				
Aesthetic appreciation	-.01	-.19	.37	.83
Inquisitiveness	-.04	-.47	.37	.63
Creativity	.03	1.16*	.46	3.19*
Unconventionality	.01	.47	.47	1.61
Altruism	-.08**	-1.13*	.60	.30*
Constant		-5.36	3.69	.01

Note. $N = 1352$. e^B = exponentiated B.

* $p < .05$. ** $p < .01$.

Table 8

Summary of logistic regression analysis for inappropriate respondents ($n = 68$) and appropriate respondents ($n = 1284$) on factor-level scores of the HEXACO-PI-R

	Bivariate correlation	B	$SE(B)$	e^B
Honesty-Humility (H)	-.02	.00	.01	1.09
Emotionality (E)	.05	.01	.01	1.00
Extraversion (X)	-.06*	-.01	.01	1.01
Agreeableness (A)	-.07**	-.02*	.01	.99*
Conscientiousness (C)	-.01	.00	.01	1.00
Openness to Experience (O)	-.01	.00	.01	1.00
Constant		.09	.06	1.00

Note. $N = 1352$. e^B = exponentiated B.

* $p < .05$. ** $p < .01$.