

# Technical documentation

## Personality test

*Bart Dekker MSc.  
Dr. Edwin van Thiel*



*Last updated on 01 April, 2020*

# Inhoudsopgave

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Method</b>	<b>3</b>
2.1	Instruments used . . . . .	3
<b>3</b>	<b>Exploration and selection of norm data</b>	<b>4</b>
3.1	Age . . . . .	4
3.2	Sex . . . . .	4
3.3	Level of education . . . . .	5
3.4	Nationality . . . . .	5
3.5	Labour market position . . . . .	6
3.6	Work sector/industry . . . . .	7
3.7	Completion time . . . . .	8
3.8	Response Variation . . . . .	8
3.9	Human response . . . . .	9
<b>4</b>	<b>Analysis</b>	<b>10</b>
4.1	Raw scores . . . . .	10
4.1.1	Factors . . . . .	10
4.1.2	Facets . . . . .	11
4.2	Correlations . . . . .	16
4.2.1	Correlations between factors . . . . .	16
4.2.2	Correlations between facets . . . . .	17
4.3	Reliability . . . . .	18
4.3.1	Factors . . . . .	18
4.3.2	Facets . . . . .	19
4.4	Construct Validity: Factor Analysis . . . . .	19
4.4.1	Screeplot . . . . .	19
4.4.2	Principal Components Analysis . . . . .	21
<b>5</b>	<b>Norms</b>	<b>22</b>
5.1	Labour force . . . . .	22
5.2	Group differences . . . . .	22
5.2.1	Sex . . . . .	23
5.2.2	Age . . . . .	24
5.2.3	Level of education . . . . .	25
<b>6</b>	<b>Conclusion</b>	<b>26</b>
<b>7</b>	<b>References</b>	<b>27</b>

## 1 Introduction

This document provides insight into the psychometry of the Big Five personality test of 123test. This test, developed by 123test B.V., is an operationalisation of the Big Five personality theory.

The test measures the five main dimensions of personality and the 30 underlying facets. This makes it a scientific instrument, which, moreover, has a high degree of reliability. has a high validity and a representative and recently assembled norm group used.

Information on reliability, validity and norm groups are described in this document. Also discussed how the dimensions of this test vary as a function of level of education, gender and age.

## 2 Method

Since November 1, 2019, more than 500,000 responses of the Big Five Personality Test have been recorded on *www.123test.com*. By analyzing the data of these anonymous respondents, we can form a good picture of this instrument.

### 2.1 Instruments used

The Big Five Personality Test is free to use at <https://www.123test.com/personality-test/>. The Dutch equivalent of this questionnaire can be found at <https://www.123test.nl/persoonlijkheidstest/> and can also be used free of charge.

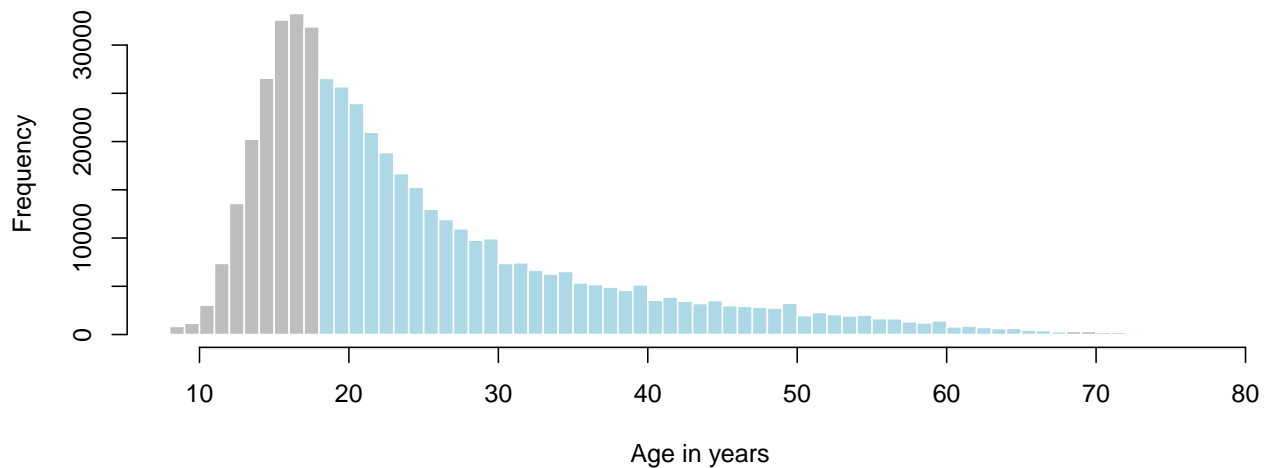
### 3 Exploration and selection of norm data

In order to explore the gathered data, this chapter examines a number of background variables of the respondent in more detail. The selection criteria used for the final dataset are indicated for each component.

The complete dataset consists of **490.689** respondents. Based on criteria such as *age*, *sex*, *educational level*, *nationality*, *labour market position*, *completion time* and *response variation*, the final subset is made on which analyses are done and with which the final norm is calculated.

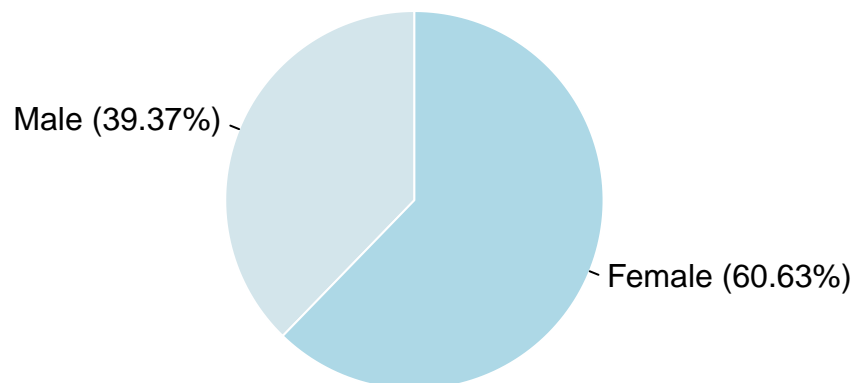
#### 3.1 Age

An age group of 18 to 67 years has been chosen, because this group best represents the working population of the Western world.



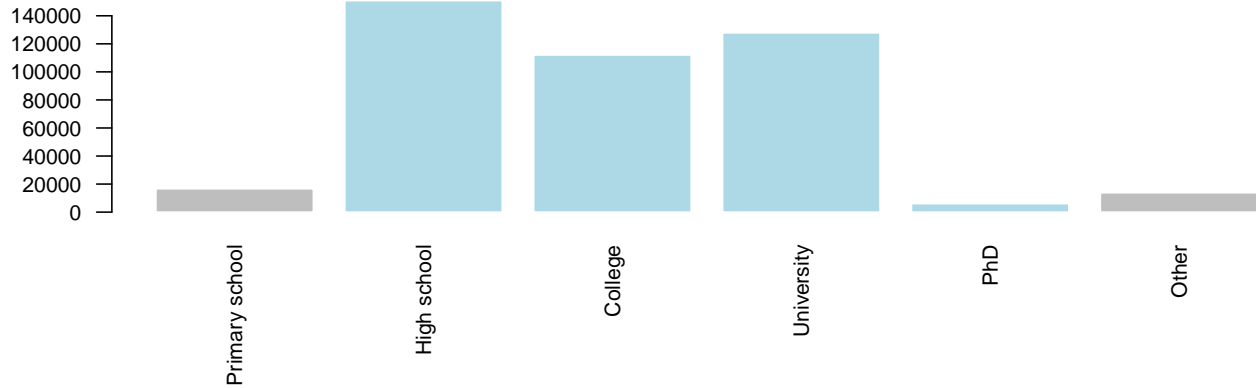
#### 3.2 Sex

The sex of all respondents is known because this background question was mandatory. Striking is the higher number of women who completed the test. Logically, both sexes are included in the dataset because this group best represents the labour population of the Western world.



### 3.3 Level of education

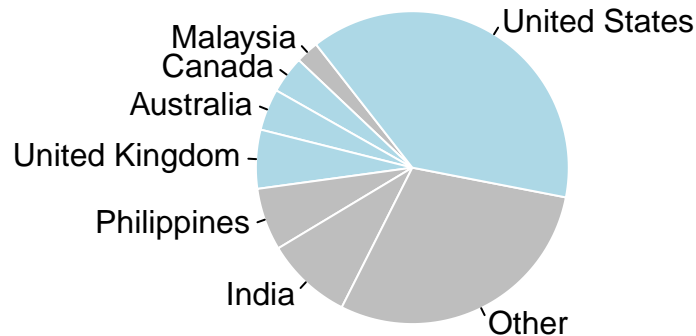
Because the Big Five Personality Test is specially developed for average to higher educated people, it was decided to select a number of education levels to be included in the dataset. The blue shaded education levels in the diagram are included in the dataset.



### 3.4 Nationality

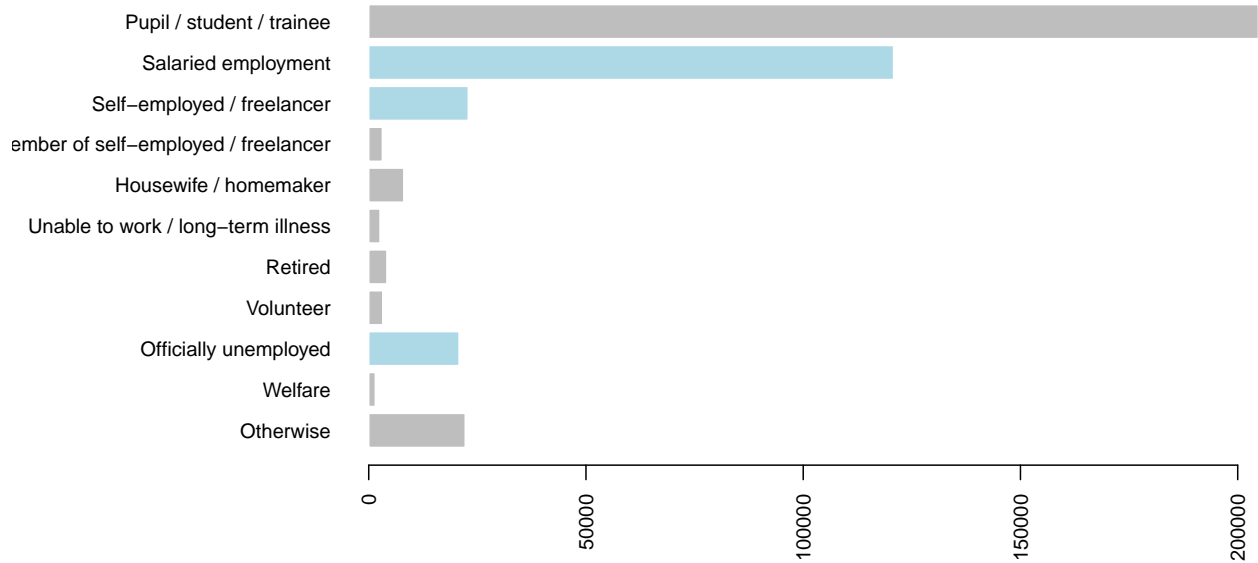
The numbers of nationalities represented in the original dataset is enormous: 217 countries, dependencies and territories were represented with more than 10 respondents.

Because the Big Five Personality Test is developed for the English speaking market of the Western world, a country selection is made. Countries selected in the final dataset are shaded blue.



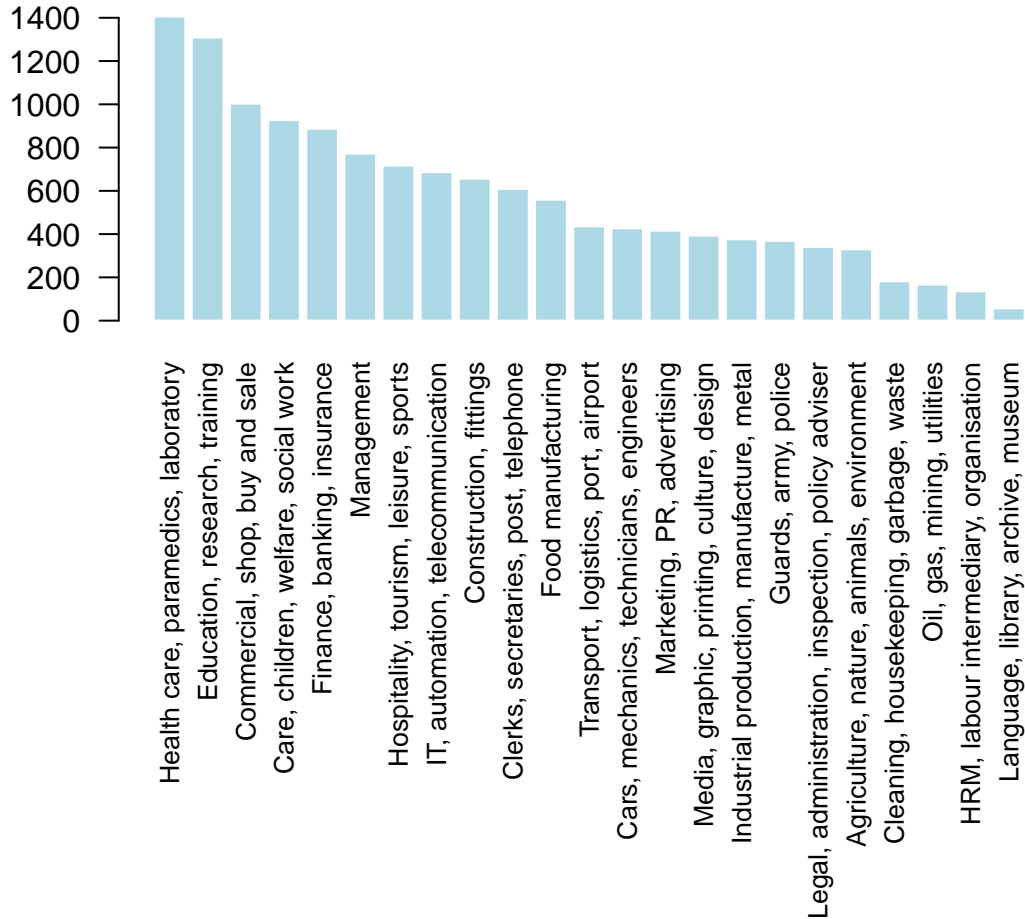
### 3.5 Labour market position

The respondent was asked about his/her labour market position. Only the labour market positions *Salaried employment*, *Self-employed/Freelancer* and *Officially unemployed* were used in the dataset, because this group best represents the labour population of the Western world.



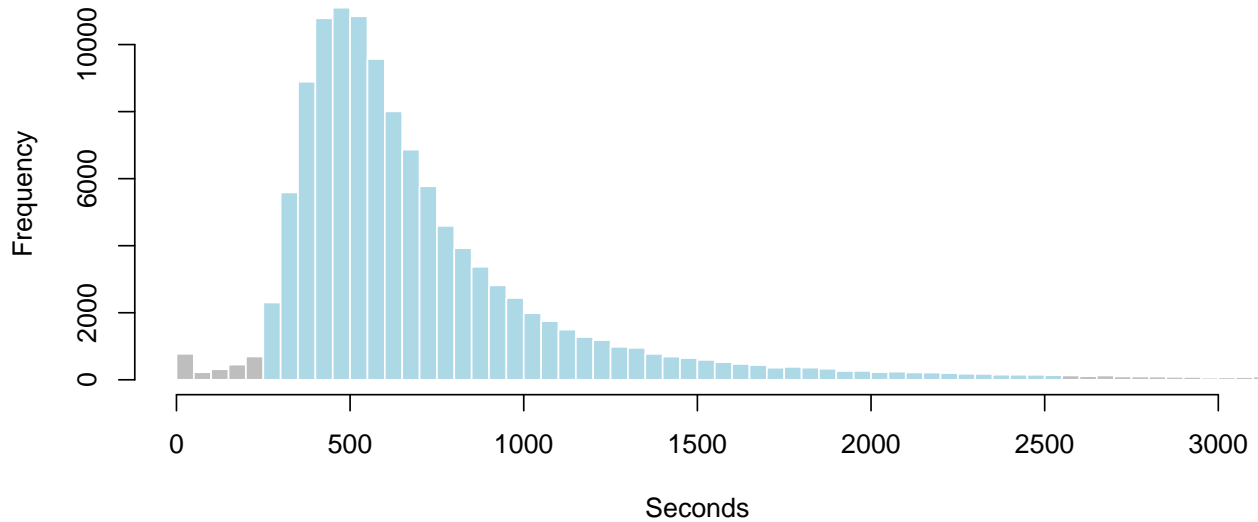
### 3.6 Work sector/industry

The respondent was asked to indicate in which working sector he/she works. A choice could be made from the 23 work sectors used in the model of *EurOccupations* (Wageindicator.org 2009). The distribution gives no reason to correct for this.



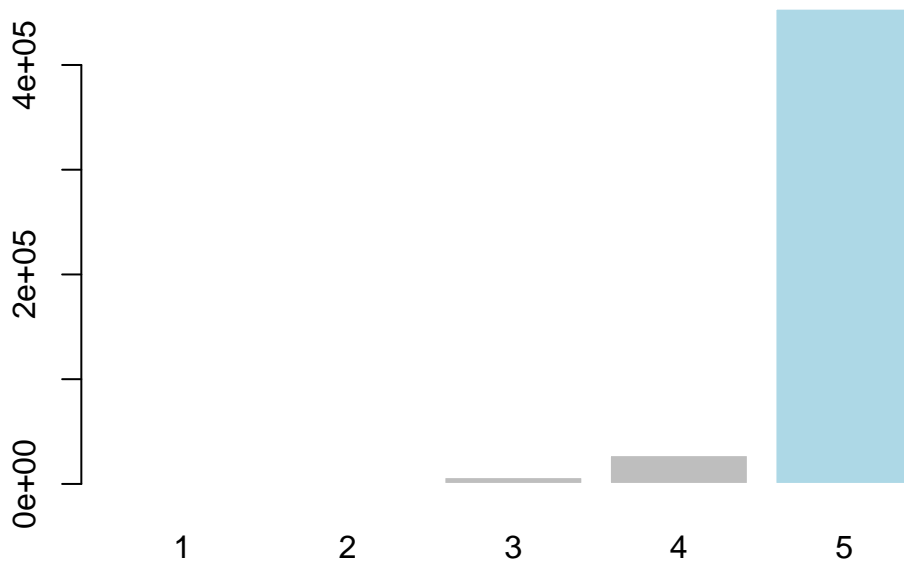
### 3.7 Completion time

Looking at the duration of completion of a questionnaire is a good way to determine how seriously a respondent has completed the questionnaire. It was decided to take between 5 and 45 minutes in the final dataset.



### 3.8 Response Variation

Looking at a respondent's response variation is a good way to determine how seriously the respondent has completed the questionnaire. It was decided to only include a response variation of 5 in the final dataset. A response variation of 5 means that a respondent has used all the answer options of the Likert-5 scale at least once over all 120 items.



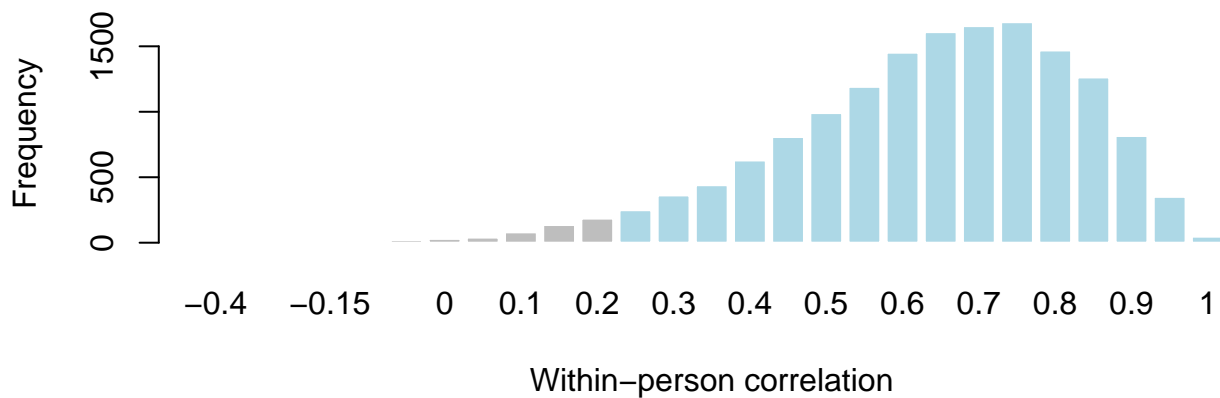


### 3.9 Human response

Online questionnaires can suffer from *crawlers* and *bots* who fill in the questionnaires automatically. By using a consistency measure we can exclude responses that are not consistent from the dataset.

The consistency measure *psychometric synonym* (Meade and Craig 2012) has been used to identify artificial and random responses. This consistency measure is calculated by first selecting all item pairs that correlate  $> .60$  across the entire dataset. In this dataset, 9 item pairs are selected. Next, for each respondent the *psychometric synonym score* is calculated which is equal to the within-person correlation of the selected item pairs.

The cut-off value of 0.2 used by Meade & Craig (2012) was used to filter artificial responses and responses with a random response pattern from the dataset. In the histogram below, the deleted responses are shaded gray.



## 4 Analysis

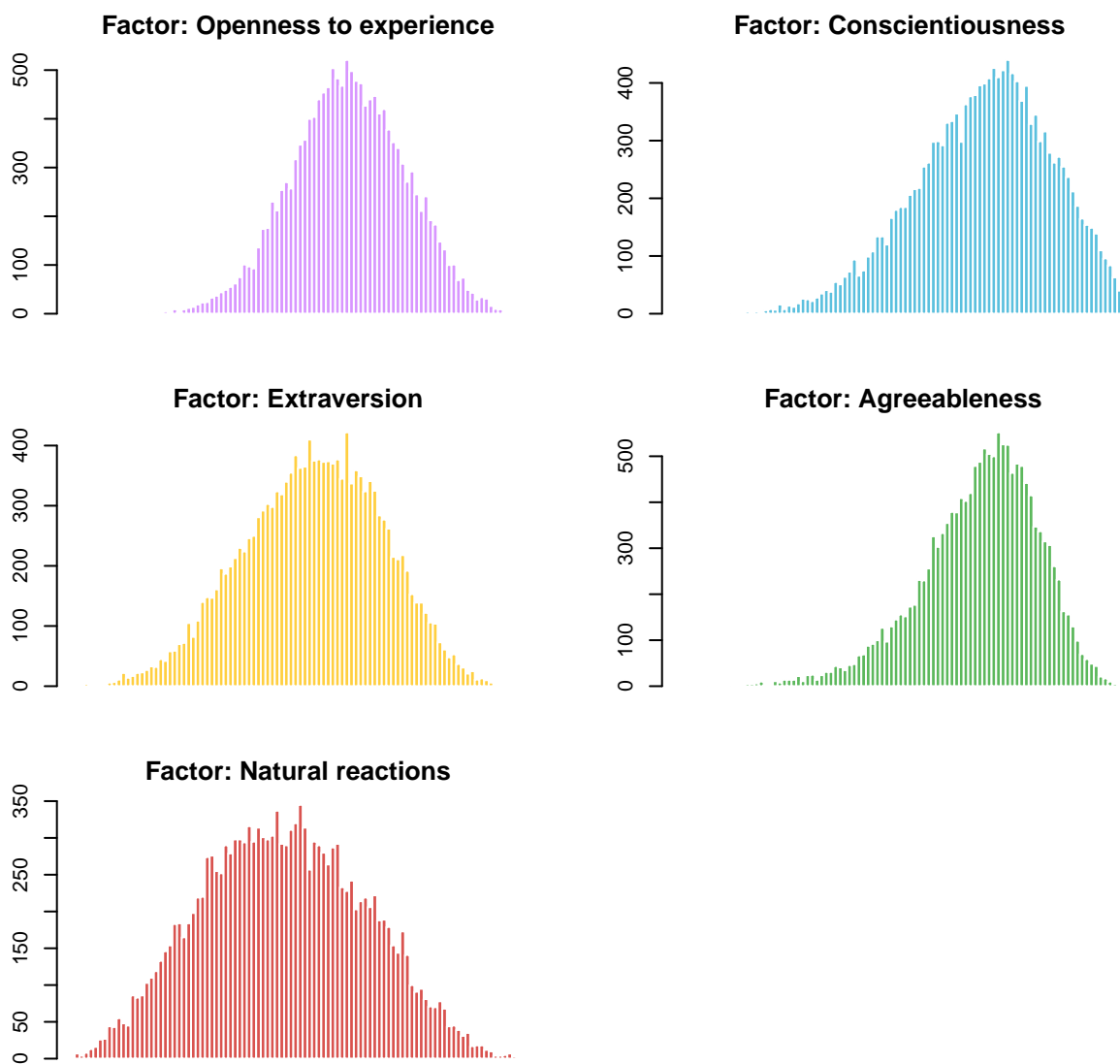
The final dataset includes *15.107* respondents.

### 4.1 Raw scores

In this chapter the raw scores of all the facets and factors are presented. X-axes have been omitted because of possible unwanted reuse of the norm data.

#### 4.1.1 Factors

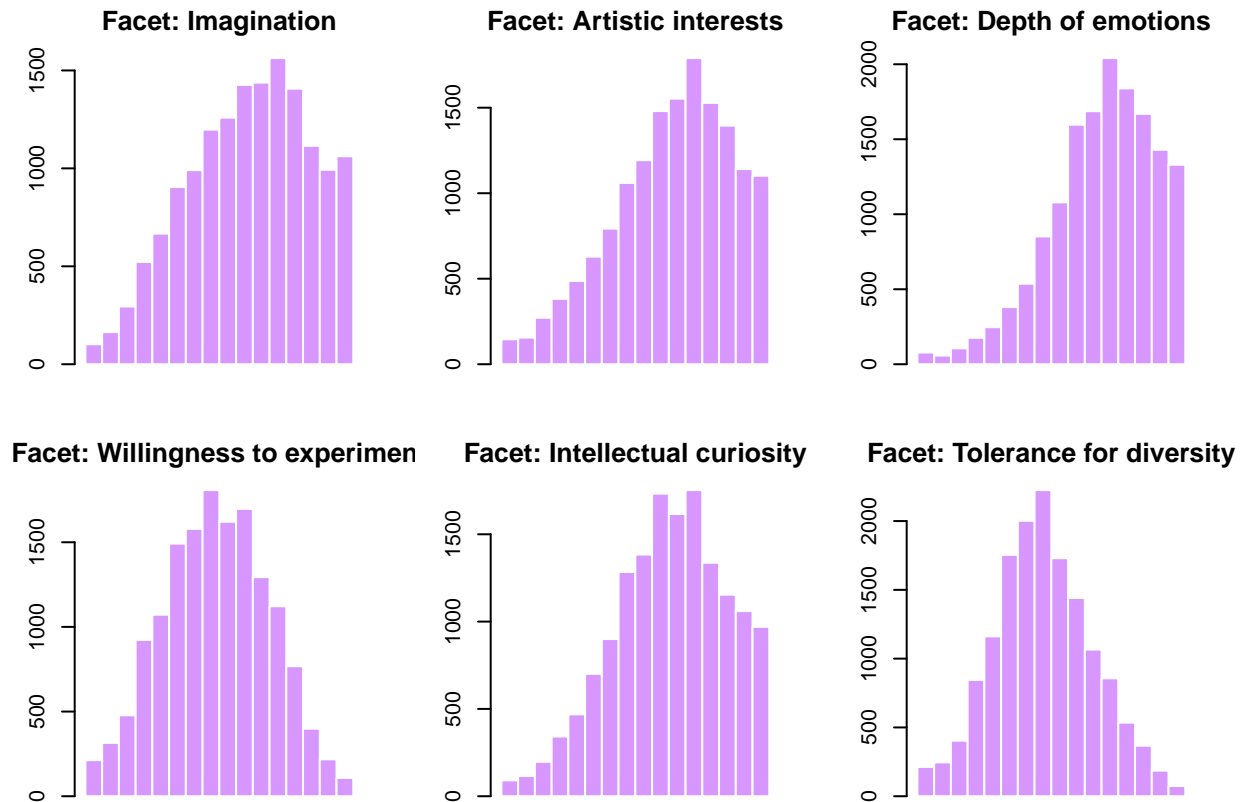
The histograms of the raw factor scores all show a normal distribution.



### 4.1.2 Facets

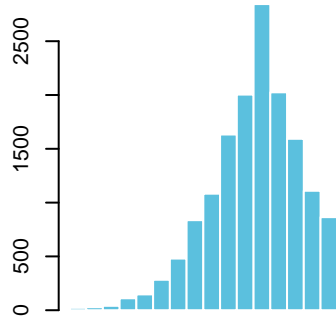
The histograms of the raw facet scores usually show a normal distribution, with sometimes a floor and/or ceiling effect for scales with a high socially desirable component.

#### 4.1.2.1 Openness to experience

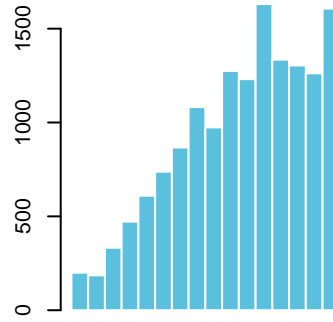


4.1.2.2 Conscientiousness

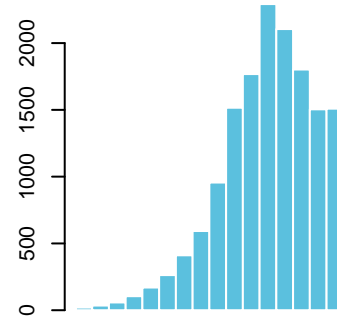
**Facet: Sense of competence**



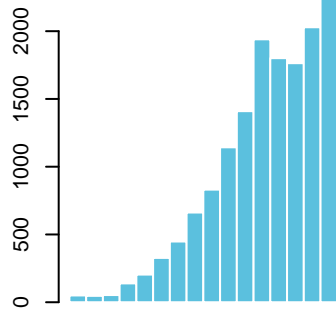
**Facet: Orderliness**



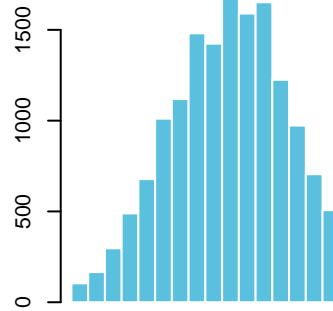
**Facet: Sense of responsibility**



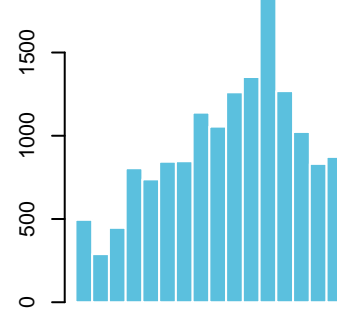
**Facet: Achievement striving**



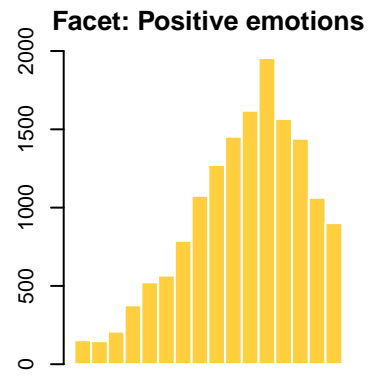
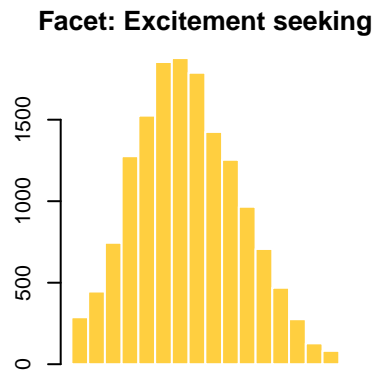
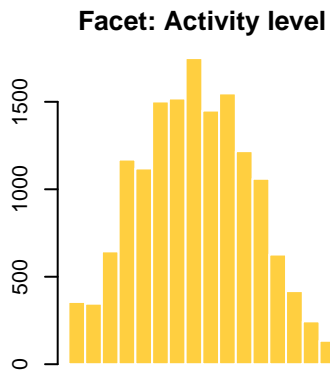
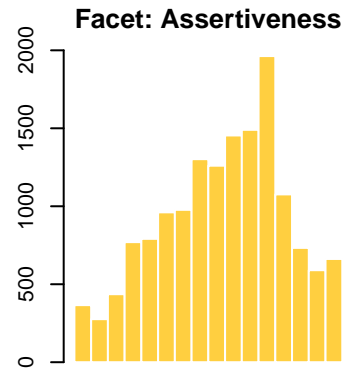
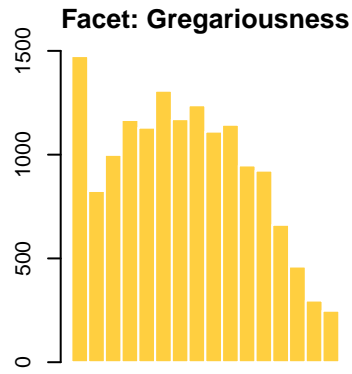
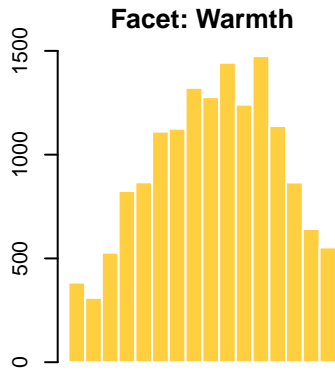
**Facet: Self-discipline**



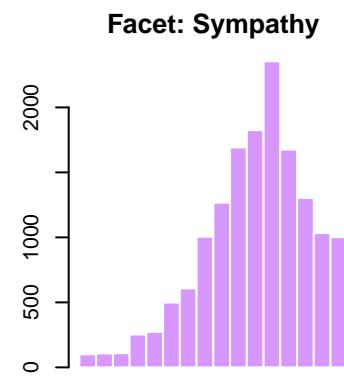
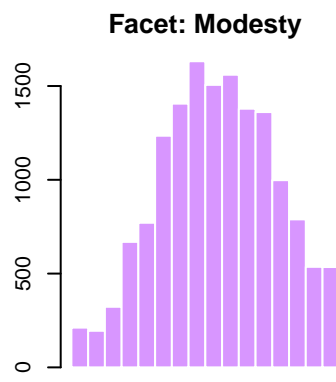
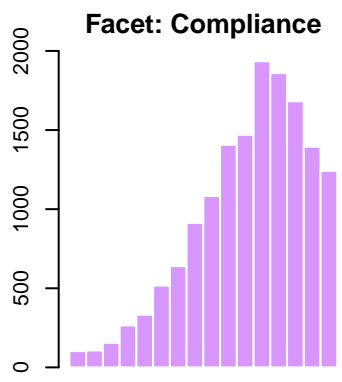
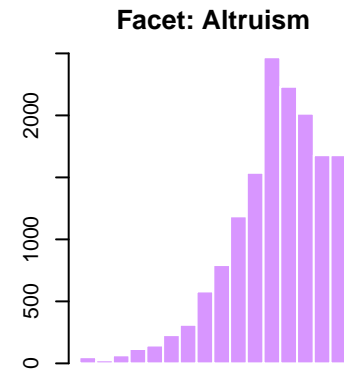
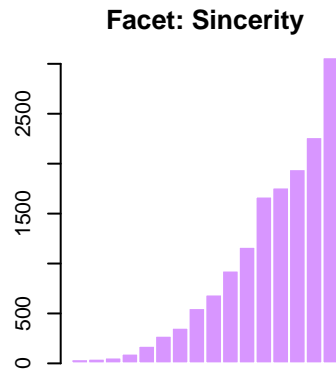
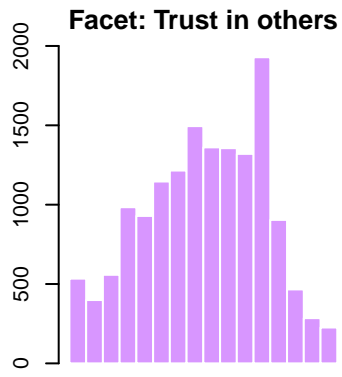
**Facet: Deliberateness**



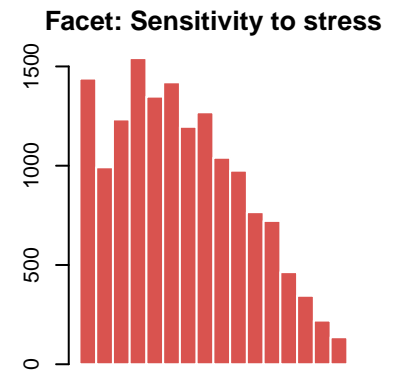
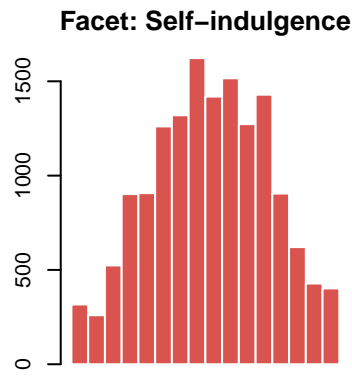
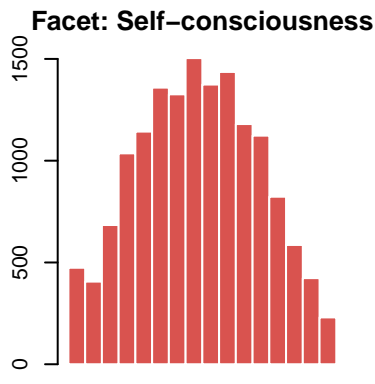
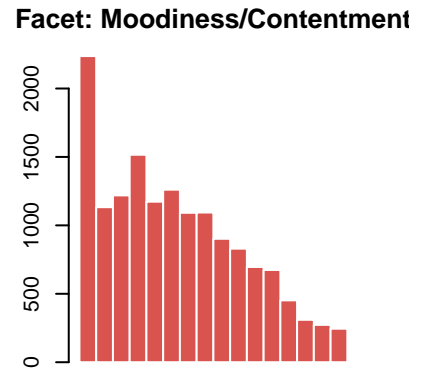
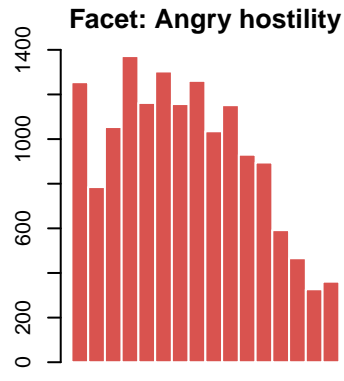
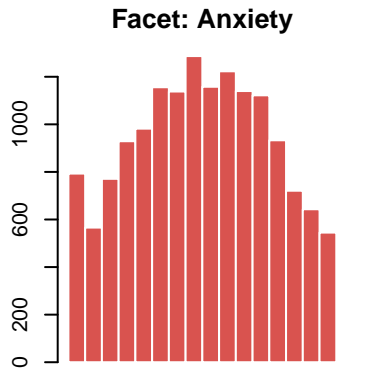
4.1.2.3 Extraversion



4.1.2.4 Agreeableness



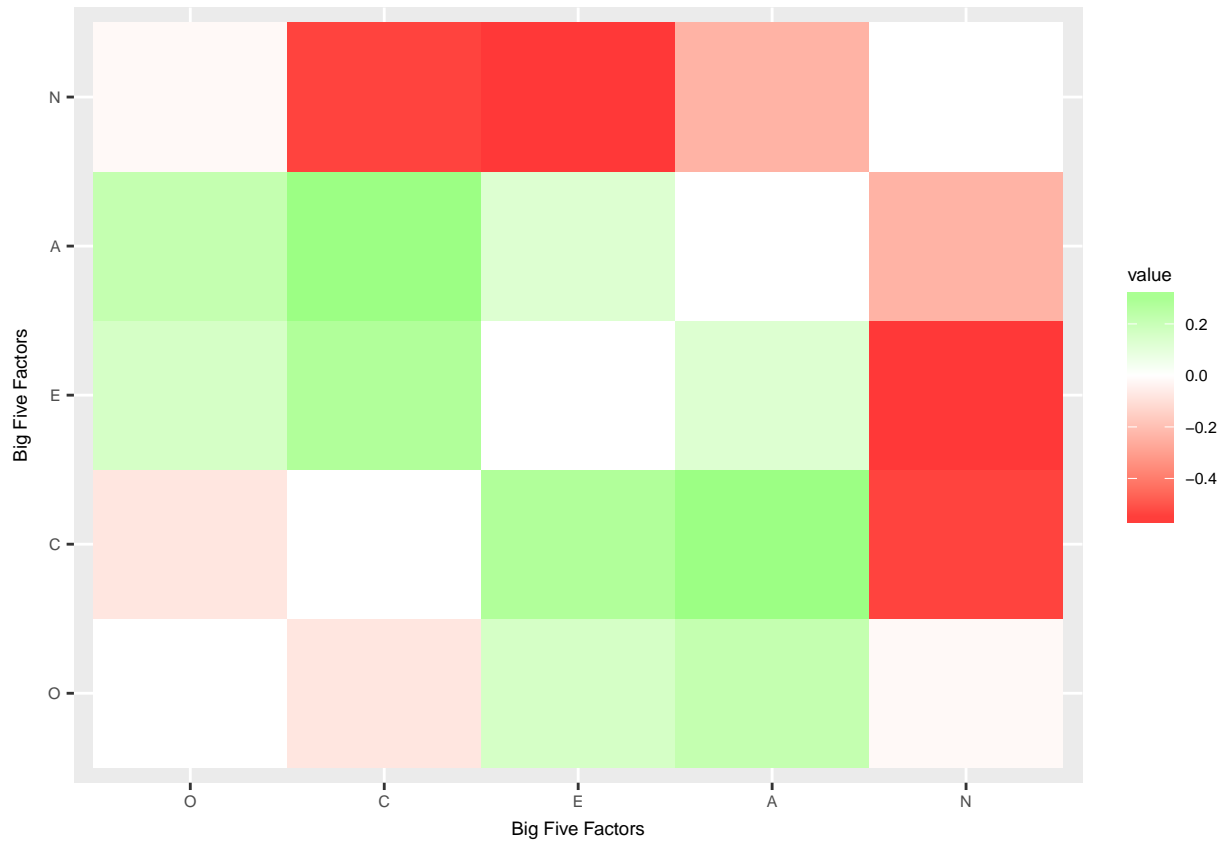
4.1.2.5 Natural reactions



## 4.2 Correlations

### 4.2.1 Correlations between factors

The five factors generally show minimal correlations. *Natural reactions* shows clear negative correlations with *Conscientiousness* and *Extraversion*.



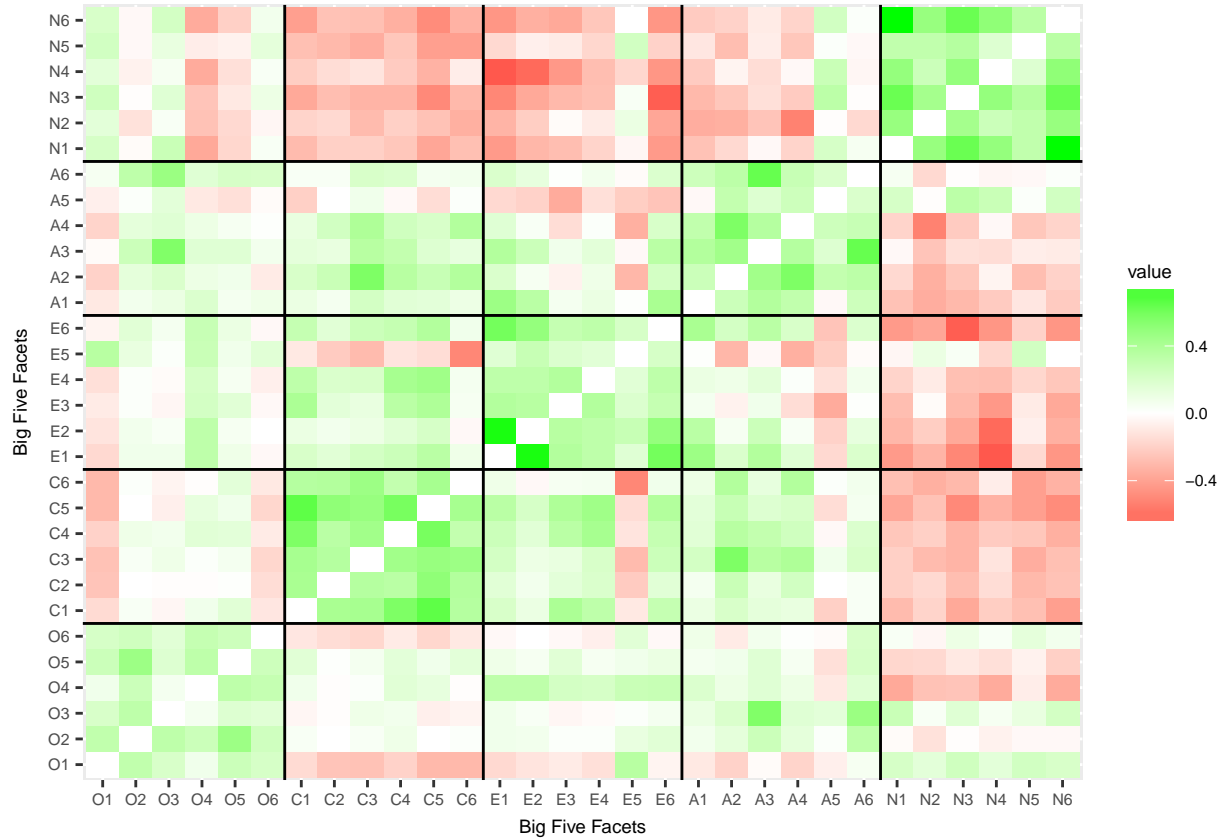


### 4.2.2 Correlations between facets

The 30 facets generally show minimal correlations outside their own facets, and average correlations within their own facets.

The previously observed negative correlation between *Extraversion* and *Natural reactions* can, at the facet level, mainly be traced back to negative correlations between *Self-consciousness* (N4) and *Warmth* (E1), and *Moodiness/Contentment* (N3) and *Positive emotions* (E6).

These results are consistent with the scientific literature and provide a good first indication of the internal structure.



### 4.3 Reliability

*Cronbach's alpha* (Cronbach and Shavelson 2004) is a measure of the reliability of psychometric tests or questionnaires. The value of *alpha* is an estimate for the lower limit of reliability of the test in question.

#### 4.3.1 Factors

A often used criterion for instruments used in advisory situations is that the reliability coefficient of Cronbach's alpha should not be lower than .60. Scores higher than .80 are assessed as 'good'.

On average across the five factors, the reliability coefficient is **0.88**, which may be considered **very high**.

Factors	Item count	Cronbach's Alpha
Openness to experience	24	0.81564
Conscientiousness	24	0.90888
Extraversion	24	0.89249
Agreeableness	24	0.86265
Natural reactions	24	0.91921

If an item does not correlate sufficiently with the other items of the same factor, it damages the reliability of said factor. Below is shown what happens to the Cronbach's Alpha of a factor when one of the 24 items is removed.

If item deleted	O	C	E	A	N
1	0.806	0.904	0.885	0.859	0.916
2	0.808	0.905	0.884	0.857	0.914
3	0.809	0.907	0.885	0.859	0.914
4	0.809	0.906	0.887	0.857	0.912
5	0.800	0.906	0.884	0.855	0.916
6	0.806	0.906	0.884	0.855	0.915
7	0.805	0.904	0.887	0.854	0.917
8	0.804	0.905	0.885	0.858	0.917
9	0.811	0.906	0.888	0.857	0.913
10	0.811	0.907	0.888	0.855	0.914
11	0.814	0.908	0.888	0.855	0.913
12	0.814	0.905	0.888	0.855	0.916
13	0.814	0.905	0.889	0.861	0.917
14	0.811	0.906	0.890	0.859	0.918
15	0.814	0.904	0.890	0.853	0.919
16	0.807	0.907	0.894	0.856	0.919
17	0.807	0.904	0.889	0.856	0.918
18	0.806	0.904	0.889	0.866	0.920
19	0.806	0.903	0.898	0.865	0.920
20	0.805	0.905	0.897	0.861	0.918
21	0.809	0.907	0.887	0.859	0.913
22	0.817	0.905	0.886	0.857	0.914
23	0.813	0.905	0.888	0.855	0.914
24	0.812	0.903	0.888	0.857	0.916

If item deleted	O	C	E	A	N
-----------------	---	---	---	---	---

### 4.3.2 Facets

The average Cronbach's Alpha of the 30 facets is **0.753**, which is a good performance considering the length of the scales.

Factors	Facets	Item count	Cronbach's Alpha
Openness to experience	Facet: Imagination	4	0.77087
	Facet: Artistic interests	4	0.73223
	Facet: Depth of emotions	4	0.65264
	Facet: Willingness to experiment	4	0.66261
	Facet: Intellectual curiosity	4	0.69076
	Facet: Tolerance for diversity	4	0.49487
Conscientiousness	Facet: Sense of competence	4	0.72658
	Facet: Orderliness	4	0.83003
	Facet: Sense of responsibility	4	0.69670
	Facet: Achievement striving	4	0.75943
	Facet: Self-discipline	4	0.74301
	Facet: Deliberateness	4	0.86425
Extraversion	Facet: Warmth	4	0.80829
	Facet: Gregariousness	4	0.81566
	Facet: Assertiveness	4	0.86924
	Facet: Activity level	4	0.71323
	Facet: Excitement seeking	4	0.65720
	Facet: Positive emotions	4	0.81739
Agreeableness	Facet: Trust in others	4	0.84768
	Facet: Sincerity	4	0.74710
	Facet: Altruism	4	0.73092
	Facet: Compliance	4	0.66146
	Facet: Modesty	4	0.73948
	Facet: Sympathy	4	0.72971
Natural reactions	Facet: Anxiety	4	0.82595
	Facet: Angry hostility	4	0.86720
	Facet: Moodiness/Contentment	4	0.85947
	Facet: Self-consciousness	4	0.70584
	Facet: Self-indulgence	4	0.76216
	Facet: Sensitivity to stress	4	0.79752

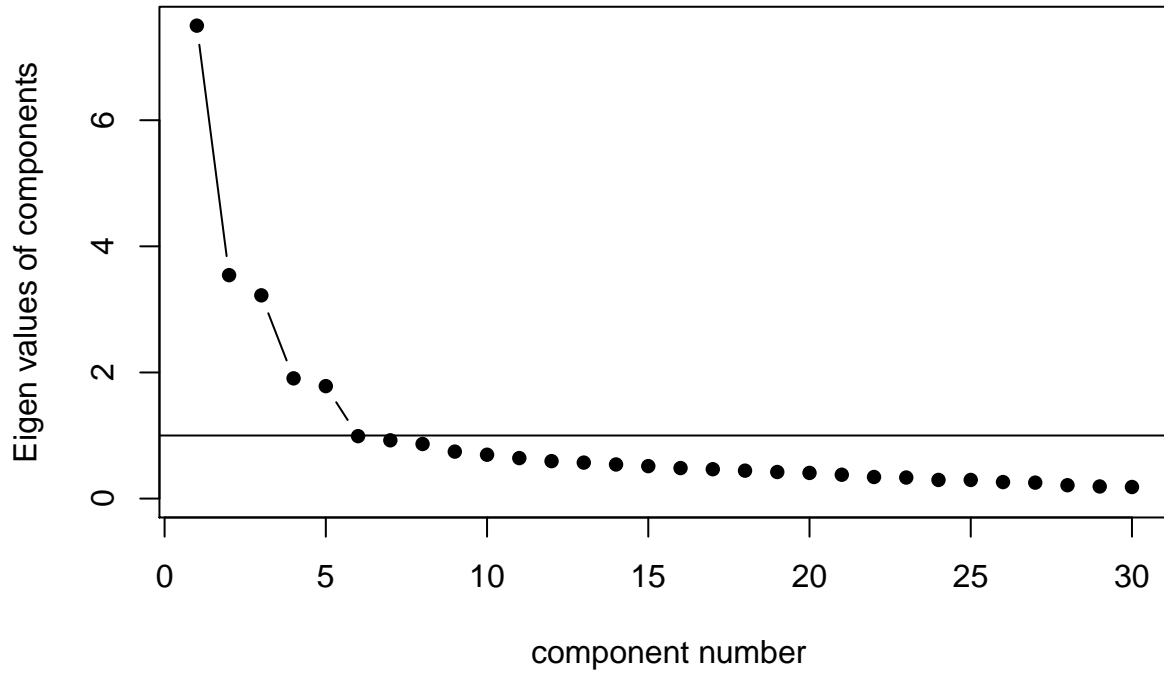
## 4.4 Construct Validity: Factor Analysis

### 4.4.1 Screeplot

In factor analysis, a *screeplot* or *eigenvalue diagram* is a graph in which the eigenvalues of the possible variables for the factors are plotted in order of decreasing magnitude.

In the table below you can see that there are 5 clear components (PC) with an eigenvalue  $> 1.0$ . This corresponds with well-known scientific literature which states that personality contains 5 components.

**Scree plot of 30 Big Five facets**



#### 4.4.2 Principal Components Analysis

*Principal component analysis* is a multivariate method of analysis in statistics to describe a large amount of data with a smaller number of relevant quantities, the main components or principal components.

The table below shows the results of a PCA with *varimax rotation*. The 30 facets can clearly be reduced to the five components to which they belong according to the theoretical model of the Big Five. The dominant factor *Extraversion* attracts a lot of variance, especially in the form of negative charges of *Natural reactions*. There are only a number of facets that have a higher primary charge on another component.

All in all, the analysis shows a very recognizable and satisfactory picture.

Factor	Code	Facet	RC1	RC2	RC3	RC4	RC5
Extraversion	E1	Facet: Warmth	<b>0.828</b>				
	E2	Facet: Gregariousness	<b>0.832</b>				
	E3	Facet: Assertiveness	<b>0.481</b>	0.513			
	E4	Facet: Activity level	<b>0.436</b>	0.493			
	E5	Facet: Excitement seeking	<b>0.574</b>				
	E6	Facet: Positive emotions	<b>0.69</b>				
Conscientiousness	C1	Facet: Sense of competence		<b>0.801</b>			
	C2	Facet: Orderliness		<b>0.578</b>			
	C3	Facet: Sense of responsibility		<b>0.529</b>			
	C4	Facet: Achievement striving		<b>0.727</b>			
	C5	Facet: Self-discipline		<b>0.784</b>			
	C6	Facet: Deliberateness		<b>0.432</b>			-0.574
Agreeableness	A1	Facet: Trust in others	0.454		<b>0.407</b>		
	A2	Facet: Sincerity			<b>0.65</b>		
	A3	Facet: Altruism			<b>0.793</b>		
	A4	Facet: Compliance			<b>0.595</b>		
	A5	Facet: Modesty			<b>0.556</b>		
	A6	Facet: Sympathy			<b>0.706</b>		
Natural reactions	N1	Facet: Anxiety				<b>0.69</b>	
	N2	Facet: Angry hostility				<b>0.708</b>	
	N3	Facet: Moodiness/Contentment				<b>0.552</b>	
	N4	Facet: Self-consciousness	-0.729				
	N5	Facet: Self-indulgence				<b>0.503</b>	
	N6	Facet: Sensitivity to stress				<b>0.638</b>	
Openness to experience	O1	Facet: Imagination					<b>0.587</b>
	O2	Facet: Artistic interests					<b>0.689</b>
	O3	Facet: Depth of emotions			0.701		
	O4	Facet: Willingness to experiment					<b>0.543</b>
	O5	Facet: Intellectual curiosity					<b>0.746</b>
	O6	Facet: Tolerance for diversity					<b>0.595</b>

## 5 Norms

### 5.1 Labour force

For a correct norm group, the dataset must properly reflect the intended group of users, in this case the Western world labour force. Because a dataset almost never has the same composition as the intended user group, *weighing* is used.

The dataset is weighted according to the distribution in the table below.

Criterion	Groepen	Population
Sexe	Female	50,50%
	Male	49,50%
Education	Average education	62,60%
	Higher education	37,40%
Age	15-24	17,40%
	25-44	45,50%
	45-64	37,10%

A much used standard for norm groups for use in ‘advisory’ situations is that the norm group should consist of >200 respondents. For recruitment and selection purposes this is >400. In this dataset **15107** respondents are included and therefore very clearly meets this standard.

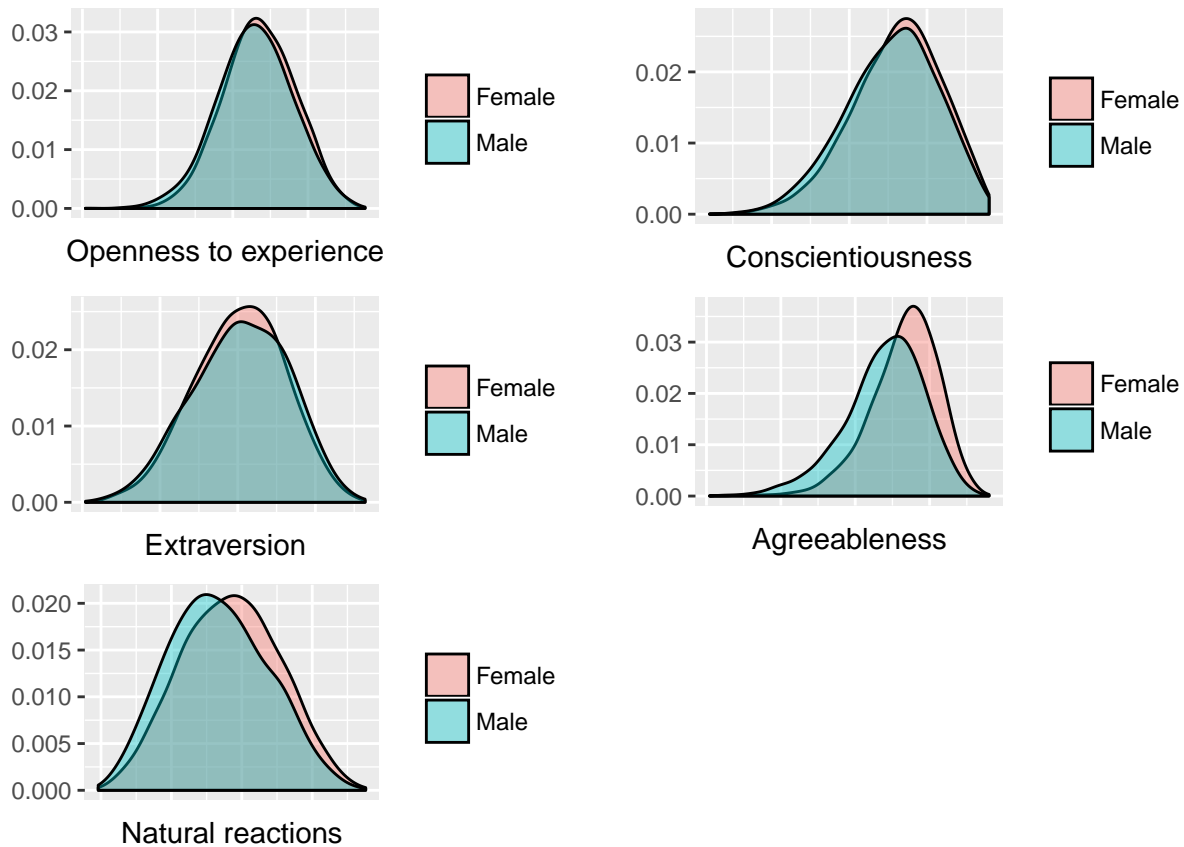
### 5.2 Group differences

If there are significant differences between relevant groups within a norm group, this could and should be corrected by using separate norm groups.

For comparing group averages and determining *effect size*, *Cohen’s d* is used (Cohen 1992). Effect sizes close to zero are small, effect sizes larger than 0.8 or smaller than -0.8 are often considered large.

### 5.2.1 Sex

To determine whether norms are needed for specific groups, group differences between the sexes have been examined. The results of these analyses are shown below.



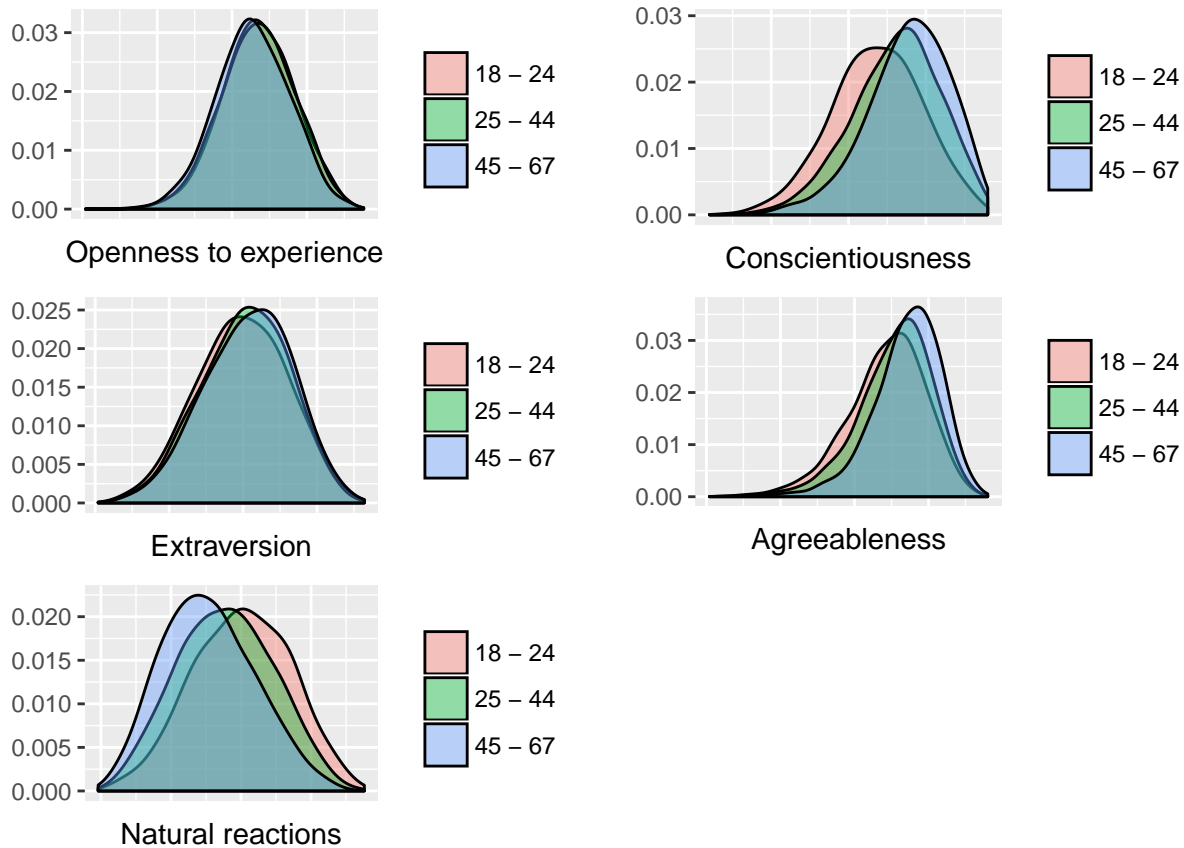
#### Effect size

Factor	Cohen's D (Male-Female)
Openness to experience	0.128
Conscientiousness	0.121
Extraversion	-0.024
Agreeableness	0.543
Natural reactions	0.252

Given that no impact sizes of -0.8 or 0.8 have been found, it can be concluded that the use of a single norm for the sexes is justified.

### 5.2.2 Age

In order to determine whether norms are needed for specific groups, group differences between age groups were taken into account. The results of these analyses are shown below.



#### Effect size

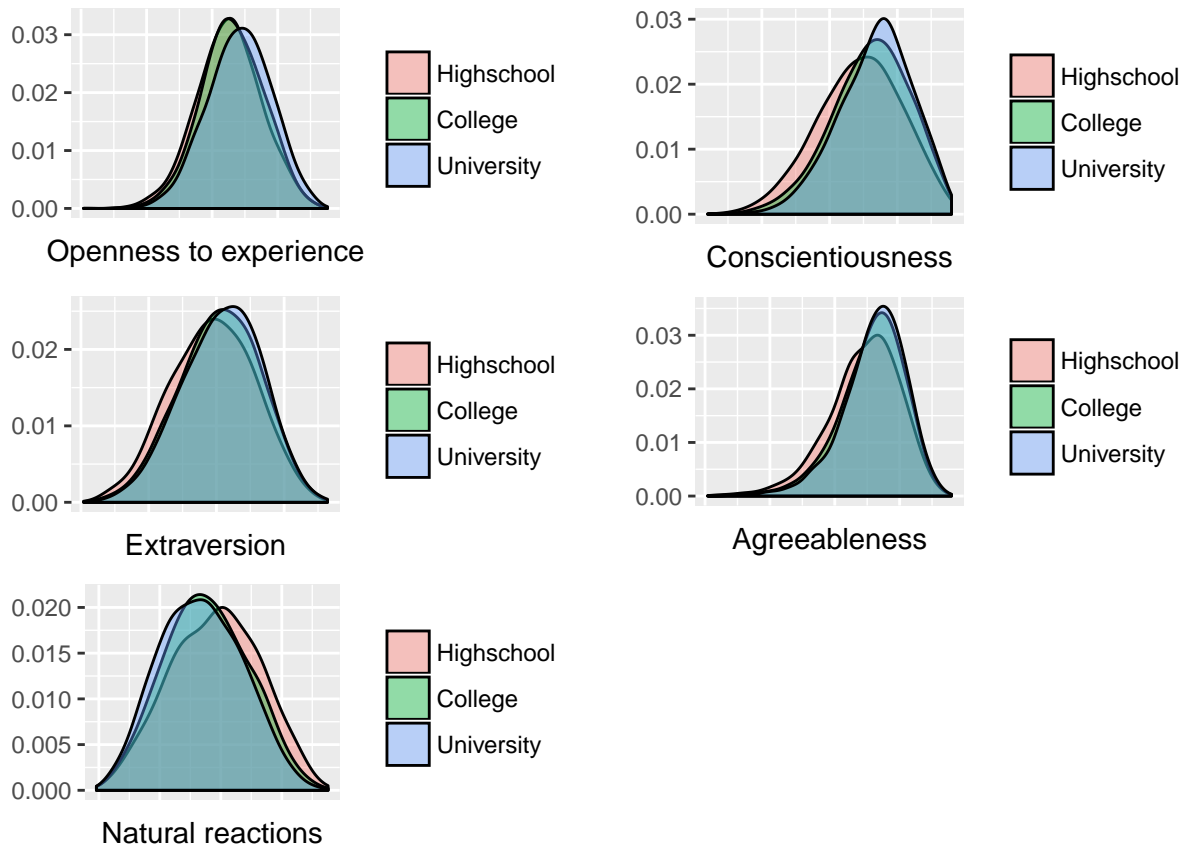
Factor	Cohen's D (Older-Younger)
Openness to experience	0.156
Conscientiousness	-0.724
Extraversion	-0.151
Agreeableness	-0.552
Natural reactions	0.705

Given that no impact sizes of -0.8 or 0.8 have been found, it can be concluded that the use of a single norm for age is justified.



### 5.2.3 Level of education

In order to determine whether standards are needed for specific groups, group differences between education levels have been examined. The results of these analyses are shown below.



#### Effect size

Factor	Cohen's D (University-Highschool)
Openness to experience	-0.330
Conscientiousness	-0.410
Extraversion	-0.242
Agreeableness	-0.312
Natural reactions	0.334

Given that no impact sizes of -0.8 or 0.8 have been found, it can be concluded that the use of a single standard for education level is justified.

## 6 Conclusion

The results of this study show that the Big Five Personality Test of 123test is a reliable and valid instrument with a solid norm to be used among Western world respondents with an average to higher educational level, with an age between 18 and 67 years for self-analysis, in career guidance or in other professional settings.

### **Reliability**

The results of this study show that the Big Five Personality Test of 123test scores well to very well on the reliability coefficients commonly used in science.

### **Validity**

The results of this study show that the Big Five Personality Test of 123test shows good construct validity of the measured constructs.

### **Norms**

The results of this study show that the Big Five Personality Test of 123test has a good norm that shows no differences between groups.

## 7 References

Cohen, Jacob. 1992. "A Power Primer." *Psychological Bulletin* 112 (1). American Psychological Association: 155.

Cronbach, Lee J., and Richard J. Shavelson. 2004. "My Current Thoughts on Coefficient Alpha and Successor Procedures." *Educational and Psychological Measurement* 64 (3): 391–418.

Meade, Adam W, and S Bartholomew Craig. 2012. "Identifying Careless Responses in Survey Data." *Psychological Methods* 17 (3). American Psychological Association: 437.

Wageindicator.org. 2009. "EurOccupations."