16PF® Questionnaire
Dutch version of the 16PF Fifth Edition

Dutch Data Supplement 2011

December 2011

unlocking potential

OPP Ltd, Elsfield Hall, 15–17 Elsfield Way, Oxford OX2 8EP

t: +44 (0)845 603 9958

f: +44 (0)1865 557483

www.opp.eu.com

Table of contents

Introduction

This 16PF data supplement for the Dutch version of the 16PF questionnaire complements the current Dutch 16PF Manual and the *2007 Data Supplement*. The results reported in the data supplement are based on data gathered from a nationally representative sample of people of working age in the Netherlands and Belgium. Where appropriate, the results obtained from this sample are compared with the findings on the Dutch standardisation sample (N=500) published in the Dutch 16PF Manual by M.T. Russell and D.L. Karol (2007). This data supplement provides further evidence of the psychometric soundness and good validity of the instrument.

Sample

The sample consisted of 1,002 respondents (501 female and 501 male.) 558 of whom completed the Dutch version of the 16PF 5th Edition questionnaire between February and March 2011 via an online data collection platform, and 444 who completed the same questionnaire in 2007. The sample was representative of the Dutch-speaking working age population. Full sample details are shown in Appendix 1.

Descriptive statistics

Statistical information (mean and standard deviation) for the raw scales for the nationally representative working age sample (N=1,002) can be found in Table 1. The table also provides the same information for the Dutch standardisation sample (N=500). The “Raw score mean difference” column shows the scale mean difference between both samples.

Independent t-tests were conducted to determine if the differences in means between both samples were significant. A significant difference ($p<0.05$) was found for only two out of the sixteen Primary Factors, Reasoning (B) and Rule Consciousness (G). However, when sample sizes are large, as is the case here, even a small difference in means is statistically significant. Statistical significance merely indicates that one can be confident that there is a difference between two samples. This does not necessarily imply that the difference is practically meaningful.

In order to determine if an observed difference is not only statistically significant but also important or meaningful, effect sizes are used. The effect size (d) was calculated by subtracting the mean for the standardisation sample (N=500) from the mean for the nationally representative working age sample (N=1,002) and dividing this by the pooled standard deviation. The effect size is a standardised value, ie all effect sizes are calculated on a common scale.

The effect size results in Table 1 show that there is a moderate effect size (in excess of the traditional 0.50 cut-off) for Reasoning (0.58). The nationally representative working age sample, on average, possesses lower Reasoning (B) abilities than the respondents in the Dutch standardisation sample.

In order to gauge how substantial the differences between both samples are, let us consider how large an effect size is required in order to affect profile interpretation. An effect size of approximately +/- 0.50 corresponds to a sten difference of approximately 1 sten. In other words, for all factors except Reasoning (B), the difference between both samples is within the standard error of measurement, which is approximately 1 sten.

The Primary Factor descriptive statistics reported for the combined nationally representative working age sample have been used to create new and updated Dutch norms for the 16PF questionnaire.

Norms

Table 2 presents norms for the 16PF, based on the data gathered for the nationally representative working age sample (N=1,002). They are generally similar, but not identical, to the norms previously generated from the Dutch standardisation sample data.

Table 1. Means, standard deviations, raw score mean differences and effect sizes for 16PF Primary Factors

		Dutch-speaking working age sample (N = 1,002)		Dutch standardisation sample (N=500)*		Raw score mean difference	Effect size (d)
	Primary Factor	Mean	Standard deviation	Mean	Standard deviation		
A	Warmth	11.43	4.31	11.66	4.56	-0.27	-0.06
B	Reasoning	8.61	3.07	10.54	3.18	-1.93	-0.58
C	Emotional Stability	14.09	4.76	14.12	4.91	-0.15	-0.03
E	Dominance	13.33	4.78	13.52	4.79	-0.36	-0.08
F	Liveliness	11.39	4.61	11.83	4.96	-0.74	-0.16
G	Rule-Consciousness	13.24	4.78	12.28	4.85	1.73	0.37
H	Social Boldness	11.79	6.60	11.64	7.16	0.38	0.06
I	Sensitivity	12.50	5.11	12.31	5.39	0.65	0.13
L	Vigilance	12.40	5.12	12.90	5.44	-0.74	-0.14
M	Abstractedness	6.93	4.85	7.40	4.79	-0.74	-0.15
N	Privateness	10.98	5.29	11.34	5.61	-0.61	-0.11
O	Apprehension	10.29	5.52	10.76	5.66	-0.66	-0.12
Q1	Openness to Change	13.30	4.60	13.29	4.68	0.12	0.03
Q2	Self-Reliance	8.66	5.26	8.12	5.75	1.01	0.19
Q3	Perfectionism	11.80	5.13	11.27	5.14	1.1	0.22
Q4	Tension	10.20	4.77	10.32	5.13	-0.24	-0.05

* From the Dutch 16PF Manual by M.T. Russell and D.L. Karol (2007)

Table 2. Dutch-speaking working age sample norms (n=1,002)

Primary Factor		Sten											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Mean**	**SD**
A	Warmth	0-2	3-4	5-6	7-8	9-11	12-13	14-15	16-17	18-19	20	11.43	4.31
B	Reasoning	0-2	3	4	5-6	7-8	9-10	11	12	13-14	15	8.61	3.07
C	Emotional Stability	0-1	2-5	6-8	9-11	12-14	15-17	18	19-20	-	-	14.09	4.76
E	Dominance	0-2	3-4	5-7	8-10	11-13	14-16	17-18	19-20	21-22	-	13.33	4.78
F	Liveliness	0-1	2-3	4-5	6-8	9-11	12-14	15-16	17-18	19-20	-	11.39	4.61
G	Rule-Consciousness	0-3	4-5	6-7	8-10	11-13	14-16	17-18	19-20	21-22	-	13.24	4.78
H	Social Boldness	-	0-1	2-3	4-7	8-12	13-16	17-19	20	21-22	-	11.79	6.60
I	Sensitivity	0-1	2-3	4-6	7-9	10-12	13-15	16-17	18-19	20-21	22-24	12.50	5.11
L	Vigilance	0-2	3-4	5-6	7-9	10-12	13-15	16-18	19-20	21-22	23-24	12.40	5.12
M	Abstractedness	-	-	0-1	2-3	4-6	7-9	10-12	13-16	17-18	19-20	6.93	4.85
N	Privateness	0-1	2	3-4	5-7	8-11	12-14	15-17	18	19-20	-	10.98	5.29
O	Apprehension	0	1-2	3-4	5-6	7-10	11-14	15-16	17-18	19-20	-	10.29	5.52
Q1	Openness to Change	0-3	4-5	6-7	8-10	11-13	14-16	17-18	19-20	21	22	13.30	4.60
Q2	Self-Reliance	-	0	1-2	3-5	6-8	9-11	12-14	15-17	18-19	20	8.66	5.26
Q3	Perfectionism	0-1	2-3	4-5	6-8	9-11	12-15	16-17	18-19	20-21	22	11.80	5.13
Q4	Tension	0-1	2	3-4	5-7	8-10	11-13	14-15	16-17	18-19	20	10.20	4.77

Reliability

Reliability gauges the consistency of test results. As a generic term, it relates to a number of different aspects of consistency. Essentially, a reliable test yields the same approximate results when administered repeatedly under similar conditions. Reliability is relevant as it describes how accurately an instrument measures the construct. It is closely related to measurement error. The higher the reliability, the smaller the band width around the observed score. It is within this band width that a person's true score is most likely to be.

The aspect of reliability addressed here is that of internal consistency, or homogeneity, of the test items, as measured by Cronbach's coefficient alpha (Cronbach, 1951). Internal consistency of the 16 factors measured by the 16PF questionnaire reflects the degree to which that set of scale items is sampling the same personality domain. In statistical terms, internal consistency reliability displays how large the intercorrelation is between the items that make up each of the 16 personality scales. Cronbach's coefficient alpha essentially calculates the average value of all possible split-half reliabilities. Internal consistency can be viewed as reliability estimated from a single test administration. As the intercorrelations among items within a scale increase, reliability of the scale itself increases. Internal consistency is lowered to the degree that items on the same scale measure different traits, or to the extent that scale items are not intercorrelated. However, it needs to be noted that – even though a high reliability coefficient is desirable – it can also lead to a scale that is too narrow in measuring a construct.

Cronbach alpha coefficients for the 16PF questionnaire were calculated based on the sample of Dutch respondents described above. Table 3 presents a comparison with the Dutch standardisation data of the coefficients for each primary scale.

Two aspects need to be taken into consideration when judging the obtained reliability coefficients.

Firstly, as mentioned earlier, reliability is dependent on the breadth of the measured construct. The 16PF Primary Factors measure distinct constructs as confirmed by factor analysis (for details see the Factor Analysis section of this data supplement). Nevertheless, when measuring the factors of personality, several behavioural preferences related to the construct are covered in the items constituting one factor. For example, the items of the factor Warmth (A) do not only cover the extent to which a person cares about others, but also how much a person is interested in spending time with others and, additionally, talking about other people's personal concerns. Making the scale more reliable by removing aspects of the constructs (ie including items about only a restricted number of behavioural preferences per scale) would mean that the construct is not accounted for in its whole range. In other words, one would increase reliability at the expense of measuring the construct adequately.

Secondly, the 16PF instrument is a personality questionnaire. Personality questionnaires are usually based on self-reports and thus measure typical behaviour. Reporting on typical behaviour rather than displayed behaviour, as is the case with ability tests, introduces more variance in the responses provided by an individual (Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001), thus lowering the reliability. In addition, when judging the reliability coefficients, the

application of the 16PF questionnaire needs to be considered. The results obtained when administering the 16PF instrument are validated in a feedback process where a practitioner discusses the profile with the respondent. Due to this process, it is possible to explore a person's true score by reflecting on the person's preferences, events that may have impacted on the responses given and by finding practical evidence. Taking this into consideration, the obtained coefficients are highly satisfactory.

The reliability coefficients can also be assessed based on the criteria for rating the technical qualities of an instrument defined by the European Federation of Psychologists' Associations (EFPA). Using EFPA's rating system, three Primary Factors achieve adequate reliability (ie coefficients between 0.60 and 0.70), eleven factors show good reliability (ie coefficients between 0.70 and 0.80) and two display excellent reliability (ie coefficient larger than 0.80). No factors yielded values below 0.60. A comparison with the reliability coefficients based on the Dutch standardisation sample shows that the differences between both samples with regard to reliability are small, thus confirming that the psychometric properties of the Dutch 16PF questionnaire are stable over time.

Table 3. Internal consistency for 16PF Primary Factors

			Dutch-speaking working age sample (N = 1,002)	Dutch standardisation sample (N=500)*
	Primary Factor	**Number of items**	**Cronbach's alpha**	**Cronbach's alpha**
A	Warmth	10	0.64	0.68
B	Reasoning	15	0.70	0.77
C	Emotional Stability	10	0.78	0.79
E	Dominance	11	0.73	0.73
F	Liveliness	10	0.72	0.77
G	Rule-Consciousness	11	0.70	0.69
H	Social Boldness	11	0.87	0.85
I	Sensitivity	12	0.69	0.72
L	Vigilance	12	0.76	0.77
M	Abstractedness	10	0.75	0.73
N	Privateness	10	0.79	0.81
O	Apprehension	10	0.80	0.80
Q1	Openness to Change	11	0.65	0.66
Q2	Self-Reliance	10	0.79	0.83
Q3	Perfectionism	11	0.72	0.72
Q4	Tension	10	0.70	0.74

* From the Dutch 16PF Manual by M.T. Russell and D.L. Karol (2007)

Primary scale factor analysis

Exploratory factor analysis is a statistical technique for discovering, within a large set of variables, a smaller set of variables that can explain much of the larger domain. Raymond Cattell's original development of the 16PF questionnaire used factor analysis to identify 16 Primary Factors. Based on the sample described above (N=1002), an analysis was conducted to examine if the same 16 Primary Factors could be replicated from the Dutch 16PF questionnaire.

The factor structure of the final set of items was examined for the sample using the procedure discussed by Conn and Rieke (1994). Items within each factor were grouped into 'parcels' based upon the strength of their correlations with items within the same scale. Hence the term 'parcels' refers to small groupings of items within a scale. For each Primary Factor, three or four items were summed within each parcel in order to achieve a parcel score. Each scale was partitioned into three to four parcels, resulting in a total of 49 parcels.

These parcels, rather than separate items, were factor analysed, because it has been shown that parcels are more reliable (Berstein & Teng, 1989; Cattell & Burdsal, 1975; Gorsuch, 1983). In addition, item responses on the 16PF instrument tend to follow a bimodel rather than normal distribution, which violates one of the statistical assumptions of normal theory factor analysis. By grouping three or four items together, the distribution better approximates a normal distribution, thus providing a better estimate of the factor structure.

In accordance with Cattell's theoretical basis for the original development of the 16PF questionnaire, an oblique rather than an orthogonal factor analysis was conducted of the parcels. Principal Axis Factoring was conducted using the statistical package SPSS. This was followed by an oblique rotational method (Promax) with the Kappa value set at 3. The Primary Factor correlation structure is reported in Table 4 with absolute loadings <0.20 excluded.

Overall, the pattern shows a very good, simple structure for the 16PF Primary Factors. All but one of the 49 parcels exhibit the highest loading onto the factor to which they were assigned. The factor loadings of the parcels onto their respective factor range from 0.29 to 0.86 (median of 0.63 and mean of 0.54); and 43 of the 49 parcels (88%) showed a loading of 0.50 or higher, thus suggesting strong links between the parcels and their assigned factor. In addition, as can be seen in Table 4, there are eight cross-loadings equal to or larger than +/-0.20. All other parcels display close-to-zero loadings onto other factors, demonstrating that these parcels represent distinct constructs that are only represented in their assigned factor, and not in the remaining factors that measure other traits.

The parcels displaying cross-loadings equal to or larger than +/-0.20 all showed higher or equal loadings onto their respective factors, with only one exception (parcel F1 loading at -0.39 on Self Reliance (Q2)). The results as a whole confirming empirically the strong conceptual links between the item parcels and their assigned factors.

In summary, 16 factors are clearly defined, corresponding to Cattell's 16 Primary Factors in the US 16PF questionnaire and many other language versions of the instrument. Such a clear Primary Factor structure of the Dutch 16PF questionnaire provides strong evidence for its construct validity.

Table 4. Rotated factor pattern loadings of 16PF Primary Factors (N=1,002; 501 females, 501 males)

	Factor															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Par-cel	Q2	H	C	I	Q4	N	Q1	Q3	L	B	M	G	A	E	O	F
A1				29									47			
A2													68			
A3													48			
B1										63						
B2										70						
B3										49						
B4										53						
C1			-86													
C2			-61													
C3			-61													
E1														29		
E2		21					20							56		
E3														71		
F1	-39															31
F2	-45															51
F3	-44															62
G1												56				
G2												53				
G3												77				
H1		81														
H2		77														
H3		66														
I1				62												
I2				68												
I3				59												
L1									75							
L2									66							
L3									64							
M1											53					
M2											59					
M3											86					
N1						82										
N2						66										
N3						71										
O1															44	
O2			24												53	
O3			35												56	
Q1_1							56									
Q1_2							53									
Q1_3							67									
Q2_1	73															
Q2_2	78															
Q2_3	66															
Q3_1								70								
Q3_2								66								
Q3_3								72								
Q4_1					67											
Q4_2					69											
Q4_3					60											

Primary Factor intercorrelations

Although the factor pattern shows that the 16PF items tend to associate with their own scale and not with others, the Primary Factor scales do clearly show a predictable pattern of intercorrelations, because the factors are oblique. Table 5 presents intercorrelations of the Primary Factor scales for the sample of Dutch and Belgian respondents (N=1,002) described above.

A comparison with the results obtained for the Dutch standardisation sample as reported in the Dutch 16PF manual shows that the differences are relatively minor (absolute difference: mean 0.00, median 0.01). This suggests that the relationships between the factors of the 16PF questionnaire are stable across different samples and over time.

Table 4. 16PF Primary Factor intercorrelations (N=1,002; 501 females, 501 males)

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3
A															
B	-10														
C	-06	08													
E	07	13	34												
F	25	-03	21	31											
G	04	-02	-05	-21	-25										
H	19	03	34	58	40	-19									
I	37	-03	-23	-26	-04	19	-16								
L	-11	-11	-33	-04	-16	-06	-20	-08							
M	-00	03	-41	-12	-03	-20	-10	20	22						
N	-26	-02	-07	-21	-26	05	-38	-02	24	02					
O	13	-03	-61	-41	-15	20	-45	29	28	32	07				
Q1	14	17	13	43	20	-14	37	01	-07	23	-24	-18			
Q2	-29	06	-23	-18	-54	04	-27	08	26	23	31	12	-07		
Q3	-02	-16	01	01	-10	30	-04	02	09	-20	03	11	-08	02	
Q4	-01	02	-45	-13	-09	-01	-24	16	25	22	10	43	-17	25	-01

Note. Decimals omitted.

Summary

The results reported in this 16PF data supplement provide strong evidence of the instrument's psychometric qualities. They demonstrate that the Dutch version of the 16PF questionnaire is a reliable and valid tool for the assessment of personality. The combination of our new 2011 findings with the results obtained from the Dutch standardisation in 2007 into an enlarged sample gives us increased confidence in the psychometric properties of the tool. Users of the Dutch version of the 16PF questionnaire can be confident that the instrument remains a reliable and valid tool for the objective assessment of personality.

References

Bernstein, I.H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence of multidimensionality due to item categorization. *Psychological Bulletin*, *105*, 467–477.

Cattell, R.B., & Burdsal, C.A. (1975). The radial parcel double factoring design: A solution to the item-vs.-parcel controversy. *Multivariate Behavioral Research*, *10*, 165–179.

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(4),* 523–562.

Conn, S.R., & Rieke, M.L. (1994). *The 16PF Fifth Edition Technical Manual.* Champaign, IL: Institute for Personality and Ability Testing, Inc.

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

EFPA (2008). EFPA review model for the description and evaluation of psychological tests. Test review form and notes for reviewers – Version 3.42. Brussels: European Federation of Psychologists' Associations.

Gorsuch, R.L. (1983). *Factor Analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Russell, M.T., & Karol, D.L. (2007). *Dutch 16PF Manual*. Champaign, IL: Institute for Personality and Ability Testing, Inc.

Appendix 1: Sample description

The sample consists of 1,002 individuals, specifically sampled to be representative of the Dutch-speaking working age population. This comprises 558 respondents who completed the Dutch version of the 16PF 5th Edition questionnaire between February and March 2011 via an online data collection platform, and 444 who completed the same questionnaire in 2007.

501 of the respondents (50%) were female and 501 (50%) were male.

The breakdown of age ranges within the sample were as follows:

Age (years)	Number	Percentage
16-20	50	5.0%
21-30	261	26.0%
31-40	261	26.0%
41-50	248	24.8%
51-60	149	14.9%
61-65	33	3.3%

A broad range of educational levels were represented.

858 (86%) of the respondents lived in the Netherlands and 144 (14%) lived in Belgium, with 820 (82%) having been born in the Netherlands, 146 (15%) having been born in Belgium. 36 (3%) of the respondents were born elsewhere.

Within the Netherlands sample, the largest single group in terms of educational level was Middelbaar bereopsonderwijs (29.0%), with the breakdown as shown below:

Netherlands Educational level	Number	Percentage
Basisonderwijs	36	4.2%
Lager beroepsonderwijs (12-tot 16-jarigen)	164	19.1%
Lager voortgezet onderwijs	71	8.3%
Middelbaar bereopsonderwijs (16-tot 18-jarigen)	249	29.0%
Hoger voortgezet onderwijs	151	17.6%
WO Bachelor	113	13.2%
WO Master	49	5.7%
PhD (ofwel Promoveren)	25	2.9%

Within the Belgian sample, the largest single group in terms of educational level was Hoger scholen onderwijs (34%), with the breakdown as shown below:

Belgium Educational level	Number	Percentage
Lager onderwijs	4	2.8%
Algemeen secundair onderwijs	26	18.1%
Technisch secundair onderwijs	43	29.9%
Hoger scholen onderwijs	49	34.0%
Universiteit – Kandituur	2	1.4%
Universiteit - Licentiaat	15	10.4%
Universiteit - Doctoraat	5	3.5%

Within the 2011 sample, respondents lived in different regions of the Netherlands, as follows:

2011 Sample		
Region	**Number**	**Percentage**
Friesland	28	5.0%
Groningen	25	4.5%
Drenthe	17	3.0%
Zeeland	17	3.0%
Noord-Brabent	76	13.6%
Limburg	39	7.0%
Noord-Holland	87	15.6%
Zuid-Holland	125	22.4%
Utrecht	37	6.6%
Gelderland	61	10.9%
Overijssel	30	5.4%
Flevoland	16	2.9%

This level of regional data analysis was not available for the 2007 sample. However, the breakdown of country residence for the 2007 sample is as follows:

2007 Sample		
Country	**Number**	**Percentage**
Belgium	144	32%
Netherlands	300	68%

The majority of the overall group were in full-time employment:

Employment status	Number	Percentage
Full-time (voor een werkgever)	549	54.8%
Self employed (zelfstandig ondernemer)	59	5.9%
Part-time (voor een werkgever)	266	26.5%
Werkloos	17	1.7%
Werk niet voor inkomen	13	1.3%
Student	33	3.3%
Huisvrouw/man	48	4.8%
Gepensioneerd	17	1.7%

Amongst the 874 people in paid employment, the occupational levels were as follows:

Occupational level	Number	Percentage
Directie/eigenaar	37	4.2%
Top management	37	4.2%
Midden management	187	21.4%
Operationeel management	79	9.0%
Werknemer	497	56.9%
Anders	37	4.2%

A range of work areas were represented. Due to the way that the separate datasets were collected, different categories were used to classify work areas for the 2007 and 2011 samples.

Within the 2011 sample, these areas were as follows:

Work area	Number	Percentage
Lansbouw, tuinbouw en visserij	4	0.7%
Mijn- en metaalindustrie	12	2.2%
Productie	33	5.9%
Energiesector (Elektra/gas)	8	1.4%
Watersector (waterschap)	1	0.2%
Bouw	8	1.4%
Groot en detailhandel	50	9.0%
Horeca (café, restaurant, hotel)	26	4.7%
Vervoer (openbaar), transport, vrachtvervoer	16	2.9%
Informatie en communicatiesector	31	5.6%
Bank en verzekeringssector	26	4.7%
Makelaardij	2	0.4%
Professionele. Wetenschappelijke en technische activiteiten	16	2.9%
Administratieve en ondersteunende diensten	39	7.0%
Overheid en Defensie	39	7.0%
Onderwijs	42	7.5%
Gezond en Welzijnszorg	92	16.5%
Kunst, Entertainment en Recreatie	13	2.3%
Dienstverlening anders	79	14.2%
Activiteiten van huishoudens als werkgever ongedifferentiee	19	3.4%
Activiteiten van extra territoriale organisaties en lichamen	2	0.4%

Within the 2007 sample, the work areas were represented as follows:

Work area	Number	Percentage
Science & Engineering	26	5.9%
Research & Development	3	0.7%
IT	61	13.7%
Health & Social Services	58	13.1%
Education	19	4.3%
Finance	22	5.0%
HR, Training & Guidance	7	1.6%
Other Business Services	24	5.4%
Administration or Secretarial	58	13.1%
Sales, Customer Service	55	12.4%
Leisure, Personal Service	3	0.7%
Military, Police, Prison Workers, Fire Brigade	11	2.5%
Land, Sea or Air Transport	9	2.0%
Skilled Operative	18	4.1%
Unskilled Operative	5	1.1%
Other private sector	3	0.7%
Other public sector	24	5.4%
Other (general)	38	8.6%