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

Data Supplement 2010

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 French version of the 16PF questionnaire complements the current *Manuel 16PF*. The results reported in the data supplement are based on actual online usage of the instrument in various HR interventions between 2003 and 2010. Where appropriate, the results obtained from this sample are compared with the findings on the French standardisation sample (N=1,000) published in the French *Manuel 16PF* by M.T. Russell and D.L. Karol (1995). This data supplement provides further evidence of the psychometric soundness and good validity of the instrument.

Sample

Originally, the sample consisted of 3,519 respondents (1,183 female and 2,336 male) who completed the French version of the 16PF 5th Edition questionnaire between 2003 and 2010 on the two online platforms 16PFWorld and OPPassessment. To ensure an equal proportion of males and females in the analyses, the sample was stratified on the basis of gender. The final sample contained 1,183 female and 1,183 male respondents (2,366 respondents overall).

Descriptive statistics

Statistical information (mean and standard deviation) for the raw scales for the overall sample (N=2,366) can be found in Table 1. The table also provides the same information for the French standardisation sample (N=1,000). 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 ($p<0.5$) was found for all Primary Factors except Perfectionism (Q3). 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=1,000) from the mean for the sample of actual usage online (N=2,366) 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) on the Primary Factors Emotional Stability (0.55), Vigilance (-0.55) and Abstractedness (-0.52). The sample containing cases of actual online usage of the 16PF questionnaire is on average more emotionally stable (C), less vigilant (L) and less theoretical (M) than the respondents in the French 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 the Primary Factors Emotional Stability (C), Vigilance (L) and Abstractedness (M) the difference between both samples is within the standard error of measurement, which is approximately 1 sten.

In addition, the differences between both samples reported here are conceptually sound and confirm what is known about personality traits and sample characteristics. Considering the typical application of the 16PF questionnaire in HR interventions, the usage sample is likely to contain a larger proportion of professionals with higher levels of education. It is therefore not surprising that this sample is on average more adaptive and emotional stable (C). Across various language versions of the 16PF questionnaire, it has been found that individuals with higher levels of Reasoning (Factor B) describe themselves as less vigilant and suspicious. As the usage sample scores more than one raw score point higher on Factor B than the French standardisation sample, it is not surprising that its mean score on Vigilance is subsequently lower in comparison. The usage sample also reports to be more practical and solution-oriented (M) than the French standardisation sample.

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

		French sample – actual usage (N = 2,366)		French standardisation sample (N=1,000)*		Raw score mean difference	
	Primary Factor	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**	**(French sample – actual usage): (French standard-isation sample)**	**Effect size (d)**
A	Warmth	15.24	3.88	14.46	4.08	0.78	0.20
B	Reasoning	9.90	2.53	8.81	2.7	1.09	0.42
C	Emotional Stability	16.72	3.17	14.78	4.23	1.94	0.55
E	Dominance	14.53	4.10	13.36	4.24	1.17	0.28
F	Liveliness	12.79	4.09	12.47	4.45	0.32	0.08
G	Rule-Consciousness	13.97	4.95	11.99	5.02	1.98	0.40
H	Social Boldness	12.93	5.61	10.93	5.76	2.00	0.35
I	Sensitivity	10.18	5.68	10.95	5.56	-0.77	-0.14
L	Vigilance	7.50	4.28	9.92	4.75	-2.42	-0.55
M	Abstractedness	4.82	4.23	7.21	5.3	-2.39	-0.52
N	Privateness	9.76	4.93	10.5	5.06	-0.74	-0.15
O	Apprehension	9.47	5.14	11.44	5.04	-1.97	-0.39
Q1	Openness to Change	21.52	4.82	20.62	4.8	0.90	0.19
Q2	Self-Reliance	5.99	4.25	7.57	5.07	-1.58	-0.35
Q3	Perfectionism	13.73	4.82	13.58	4.96	0.15	0.03
Q4	Tension	7.34	4.85	9.18	5.06	-1.84	-0.37

* From the French *Manuel 16PF* by M.T. Russell and D.L. Karol (1995)

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 French respondents described above. Table 2 presents a comparison with the French standardisation data of the coefficients for each primary scale and Impression Management.

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 the factor analysis; for details see later in 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 as well as in 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, six Primary Factors achieve adequate reliability (ie coefficients between 0.60 and 0.70), seven factors show good reliability (ie coefficients between 0.70 and 0.80) and one displays excellent reliability (ie coefficient larger than 0.80). Only two factors (Factor A and Factor B) yielded values below 0.60. A comparison with the reliability coefficients based on the French standardisation sample shows that the differences between both samples with regard to reliability are negligible, thus confirming that the psychometric properties of the French 16PF questionnaire are stable over time.

Table 2. Internal consistency for 16PF Primary Factors

			French sample – actual usage (N = 2,366)	French standardisation sample (N=1,000)*
	16PF factor	**Number of items**	**Cronbach's alpha**	**Cronbach's alpha**
A	Warmth	11	0.55	0.57
B	Reasoning	15	0.58	0.59
C	Emotional Stability	10	0.65	0.70
E	Dominance	10	0.68	0.65
F	Liveliness	10	0.63	0.68
G	Rule-Consciousness	11	0.68	0.68
H	Social Boldness	10	0.83	0.81
I	Sensitivity	11	0.76	0.75
L	Vigilance	10	0.71	0.74
M	Abstractedness	11	0.71	0.76
N	Privateness	10	0.75	0.76
O	Apprehension	10	0.73	0.72
Q1	Openness to Change	14	0.67	0.62
Q2	Self-Reliance	10	0.69	0.76
Q3	Perfectionism	10	0.75	0.76
Q4	Tension	10	0.72	0.72
IM	Impression Management	12	0.59	0.58

* From the French *Manuel 16PF* by M.T. Russell and D.L. Karol (1995)

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=2,366), an analysis was conducted to examine if the same 16 Primary Factors could be replicated from the French 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 3 with absolute loadings <0.20 excluded.

Overall, the pattern shows a very good, simple structure for the 16PF Primary Factors. All 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 0.36 to 0.79 (median of 0.61 and mean of 0.62); and 45 of the 49 parcels (92%) 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 3, there are only two 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 two parcels displaying slightly higher loadings (A1 and F2) had loadings of -0.22 onto Privateness (-) and -0.20 onto Sensitivity (-) respectively. In comparison, both parcels still show considerably higher loadings onto their respective factors (0.46 and 0.53 respectively), thus confirming empirically their strong conceptual link with their assigned factor.

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 French 16PF questionnaire provides excellent evidence for its construct validity.

Table 3. Rotated factor pattern loadings of 16PF Primary Factors (N=2,366; 1,183 males, 1,183 females)

	Factor															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Par-cel	N	I	O	Q1	Q4	H	Q3	M	Q2	G	L	F	B	E	C	A
A1	-22															46
A2																37
A3																57
B1													58			
B2													62			
B3													54			
C1															36	
C2															53	
C3															67	
E1														59		
E2														64		
E3														39		
F1												59				
F2		-20										53				
F3												61				
G1										61						
G2										69						
G3										56						
H1						79										
H2						60										
H3						73										
I1		76														
I2		67														
I3		73														
L1											60					
L2											70					
L3											61					
M1								60								
M2								64								
M3								61								
N1	66															
N2	79															
N3	76															
O1			73													
O2			67													
O3			58													
Q11				61												
Q12				58												
Q13				52												
Q14				50												
Q21									64							
Q22									63							
Q23									59							
Q31							64									
Q32							79									
Q33							58									
Q41					60											
Q42					62											
Q43					71											

Note. Decimals omitted. Factor loading less than absolute value 0.2 deleted.

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 evince a predictable pattern of intercorrelations, because the factors are oblique. Table 4 presents intercorrelations of the Primary Factor scales for the sample of French respondents (N=2,366) described above.

A comparison with the results obtained for the French standardisation sample as reported in the French *Manuel 16PF* shows that the differences are relatively minor (absolute difference: mean 0.05, median 0.04). 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=2,366; 1,183 males, 1,183 females)

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3
A															
B	01														
C	12	14													
E	02	11	34												
F	34	07	18	14											
G	-05	-08	10	08	-16										
H	29	09	36	43	38	02									
I	29	03	-14	-30	10	-24	-11								
L	-18	-25	-23	-02	-15	03	-19	-13							
M	-05	-03	-35	-16	02	-33	-13	22	17						
N	-40	-08	-06	-09	-27	09	-34	-09	27	-02					
O	06	-02	-42	-32	-10	03	-33	27	17	24	-02				
Q1	13	20	23	31	25	-21	31	07	-15	14	-14	-08			
Q2	-21	-04	-17	-12	-38	-12	-26	14	20	22	23	09	-08		
Q3	-09	-19	12	19	-08	45	07	-26	20	-31	10	-01	-06	-04	
Q4	-05	02	-33	-04	-05	-21	-18	21	12	22	-03	31	-10	23	-21

Note. Decimals omitted.

Summary

The results reported in this 16PF data supplement provide excellent evidence of the instrument's psychometric qualities. They demonstrate that the French version of the 16PF questionnaire is a reliable and valid tool for the assessment of personality. A comparison of our current findings with the results obtained from the French standardisation in 1995 shows compelling similarity in the psychometric properties between the data from the two different points in time. This means that what we knew then about the French version of the 16PF questionnaire is concurrent with what we found in our more recent data. We can therefore conclude that the 16PF instrument is just as current now as it was when the data for the standardisation was gathered in 1995. Users of the French 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. (1995). *Manuel 16PF5*. Champaign, IL: Institute for Personality and Ability Testing, Inc.