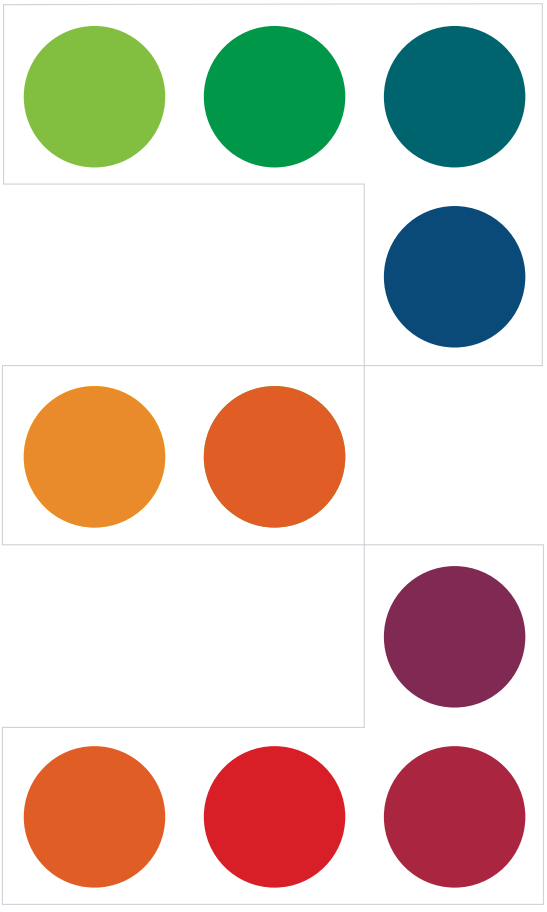
Technical summary



Contents

Norman L Buckley BSc(Psych)(Hons) CPsychol
Rebekah J Williams BA(Hons)(Psych) MSC(BehMed) MSC(OrgPsych) CPsychol

Introduction

In the mid 1980s in the UK there was a lot of discussion in the professional press about the problems associated with 'ipsative' tests. Ipsative tests are those where you are provided with different statements or adjectives, and are then required to nominate which is 'most like you' and which is 'least like you'. In some instances you might think that all the adjectives offered are good descriptors of your personality; in other instances you might think that none of the adjectives are relevant to your personality. Regardless, you are always forced to choose two of the four: one that is most like you, and one that is least like you. The issues around ipsative tests were brought into focus by an article in 1988 title 'Spuriouser and spuriouser: The use of ipsative personality tests' by Johnson et al[1].

As a result, ipsative tests can be annoying for organisational members to complete: there is a strong feeling that the test is not capturing the personality or working style of the individual in question. Moreover, and perhaps more seriously, statistical analysis of the data that emerges from ipsative testing is highly contentious. At this time I was asked to explain ipsative testing and its validity to the assembled HR managers of a large conglomerate organisation. I put my thoughts together and trundled off to a training centre (under the F111 take off path incidentally) at Upper Heyford in Oxfordshire, UK.

I was welcomed by the group HR director who ushered me in, thanked me for coming to talk about this complicated topic, and then whispered that: 'the bloke in the corner uses one of the best known ipsative tests around, so can you sort him out.' I thanked him for the hidden agenda, gave my talk in what I hoped was a suitably professional way and then stopped for questions. The F111s continued overhead. Our ipsative-using friend in the corner opened with three questions:

- Does ipsative testing tell me anything at all?
  Well yes I replied. Not sure what or how reliably but it probably tells you something.
- Do I cause any damage using it?
  Well probably not, depending on how you use it. But don't try and compare one person with another.
- What have you got, as a representative of Organisational Psychology; that is as quick, easy to use and gives me more?
  Umm – nothing. All the 'proper' tests around (16PF, MMPI, CPI etc) require a high level of training and experience to work them properly.

A perfectly timed pause, followed by: 'Then why don't you shut up, go away and come back when you have something?' I saw this comment as a design brief. From it I recognised the need for something that had both the utility of the quick tests, as well as the statistical integrity of the established psychometric heavyweights: Cattell's Sixteen Personality Factors Questionnaire (the 16PF), the Minnesota Multiphasic Personality Inventory (MMPI) and the California Psychological Inventory (CPI). This was the impetus for Facet5.

Drivers for the development of Facet5

Facet5 was designed in response to requests from organisations for a psychometric instrument that combined the utility of widely used management tools and the psychometric integrity of the mainstream personality questionnaires. The information we received centred on the following themes:

- For psychology to have business-value it needs to provide useful information: organisations require practical models of personality as well as clearly outlined conceptual frameworks.
- Currently there are multiple models and frameworks describing personality. Sometimes the same models are labelled differently; this creates confusion.
- Some models are based on Type theory, some on Trait theory. Trait theorists differ in their interpretations even when they used the same methodologies.
- Many of the team and leadership tools lack psychometric integrity.
- Psychologists in some countries 'ring fence' their knowledge base and use complicated psychological terminology to make this knowledge base hard to penetrate. It seems that the barrier to entry is a professional label rather than professional competence.

This understanding drove the development of Facet5, and the information we had received from organisations regarding the use value of psychometric testing prompted us to develop certain guidelines:

- Facet5 should to be quick. Reliable professional psychometrics at the time could take over an hour. They then required painstaking feedback from highly experienced practitioners to ensure that the correct messages were provided. Team-based and leadership-based tools could boast of completion times around the 10-minute mark. We would aim for something in between.
- Facet5 should to be easy to understand and interpret. Many (otherwise reliable) tools seemed to lose themselves in large quantities of dense text. We needed to make the results of personality tests clear to people who were not trained psychologists. We would use a graphics-based output to facilitate this aim.
- Facet5 should use natural language – not complex psychological jargon. We would use the most user-friendly language possible to communicate personality theory, measurement, results, interpretation and most importantly, organisational implications.
- Facet5 should provide 'information' not 'data.' Facet5 would use the latest psychometric testing methods to extract personality data, and would then translate this data into information that an organisation could both understand and benefit from.

Questionnaire origins, content and structure

Sources of original thinking

During the time of Facet5's development, there was an emerging consensus among organisational psychologists of a 'Big-5' model of personality, wherein personality consisted of five core 'factors'. However, the psychological literature differed on which factors should be included.
To resolve this we conducted a systematic literature review, as well as an in-depth analysis of the validity of existing personality questionnaires, including:

- The Humm-Wadsworth Temperament Scale (the first 'criterion keyed' personality questionnaire and fore-runner of the California Personality Inventory and MMPI)[2]
- Gough's California Personality Inventory (CPI)[3]
- Hathaway and McKinley's MMPI[4]
- Catell's 16PF[5]
- Eysenck's EPQ[6]
- The Myers-Briggs Type Indicator[7]
- Marston's DISC[8]

In addition a database of over 200 items generated by repertory grid studies with managers in a range of British industries were sifted for content and modified for inclusion where appropriate. These items had been created using a triadic Repertory Grid method where the topic was job performance: The question was:

'In what way are two of these people similar in terms of job performance and different from the third?'

Organisations supporting these studies included:

- An international airline where five hundred cabin-crew were rated using a complex paired comparison process, which produced a sparse matrix. This matrix was then analysed to give a rank order for all participants. We then separated the repertory grid constructs for the top ranked crew from the constructs for the bottom ranked crew;
- A global cosmetics company where we measured the relationship between behaviours extracted from repertory grids and performance figures for Area Sales Managers;
- A major tobacco company where we compared mangers' ratings of regional sales managers with the extracted behaviour constructs of those regional sales managers;
- A construction materials company where the target group were graduates and the criterion was successful completion of a graduate assessment centre.
- A major pharmaceuticals company where the target group were graduates and the criterion was again successful completion of a graduate assessment centre.

- A global consumer products company where we analysed the correlation between certain graduate behaviours and successful completion of assessment centres.

Item writing

The net result of these triadic repertory grid studies was development of specific items related to job performance. Item definition, item writing and item refinement was carried out by a team of professional psychologists including:

- academic psychologists from the University of Edinburgh
- occupational psychologists working commercially in the UK
- cognitive psychologists

Guidelines for the item writers included:

- Items need to reflect the current understanding of the personality construct in question
- Items should use commonly accepted terminology and everyday language
- Items should avoid slang or idiom (specifically to reduce culture bias)
- Items should be gender neutral
- Antonymous pairs should be value neutral
- If value neutrality is not possible antonymous pairs should be semantically balanced

The finished product of this effort was a 150-item set upon which there was general agreement. Each of the Big-5 domains was equally captured in the item set.

The item set was subsequently sent to an advisory team of HR professionals (including HR managers, recruitment consultants, HRD specialists and OD specialists from a number of different organisations within the UK). In order to determine construct validity, the advisory team was provided with a set of personality domain definitions and asked to match each item to the domain they felt the item related to. The result of this filtering was the screening out of 44 items that appeared to conflate personality domains. This left us with a 106-item set that was ready for trial.

Item format

We chose to use a semantic differential approach as opposed to the more common Likert scale. The concept of a Semantic Differential is usually credited to Charles Osgood in his defining work *The measurement of meaning*[9]. One reason for choosing this format is speed. When people make judgements (particularly about personality or similar attributes), they often use what may be termed an 'implicit opposite' to help them define a scale on which to organise their judgement. For example, if trying to rate how 'friendly' someone is, a respondent may think of the term 'hostile' as occupying the opposite end of the scale. A Semantic Differential provides both anchors for the scale whereas a traditional Likert scale provides one anchor and leaves it up to the respondent to create the opposite anchor. It has been suggested that anchoring both ends of the scale makes

judgements both easier and faster[10]. The repertory grid process helps here because it identifies antonymous adjectives, phrases or statements and places them at opposite ends of a scale. For example when asked 'In what way are two of these people similar and different from the third?' the respondents may say 'this person is organised whereas these two are disorganised'. Organised and disorganised are therefore scale anchors. Facet5 takes this style.

Administration method

The presentation process for Facet5 has changed. It was always designed to be unproctored and untimed. The initial Facet5 questionnaire layout saw the items arranged into semantic differentials with a five-point[11] scale between them[12]. In the original paper-and-pencil version, the questionnaire was presented with approximately 20 items per page and looked like this:

Facet 5 Questionaire

21	It is important for a manager to fit in with what the team wants	1 2 3 4 5	It is important for a manager to set very clear goals for others
22	Everybody has the right to a decent job	1 2 3 4 5	Many unemployed are unemployable
23	I like companies where there is an established structure	1 2 3 4 5	I don't like the idea of hierarchies in companies
24	I love a job where I have to learn the skills as I go along	1 2 3 4 5	I prefer a job where I know I have the skills to perform well
25	I believe in telling people exactly how I feel	1 2 3 4 5	I tend not to show my true feelings
26	I often join in an argument to make my point	1 2 3 4 5	I can't bear arguments for any reason
27	In most companies politics are as important as results	1 2 3 4 5	Most companies reward achievement
28	I like to have time to prepare myself for a group presentation	1 2 3 4 5	I can give group presentations without time spent preparing
29	There's no point in doing more than is necessary	1 2 3 4 5	Perfection is always worth striving for
30	I am a quiet and reserved person	1 2 3 4 5	I enjoy social activities
31	It is always possible to sort things out without an argument	1 2 3 4 5	Some people only understand when you get aggressive
32	I just take things as they come	1 2 3 4 5	As an important event gets closer I tend to worry more about it
33	In business you must obey a code of ethics	1 2 3 4 5	'All's fair in love and war'
34	At work I believe in taking things seriously	1 2 3 4 5	Work doesn't have to be a solemn affair
35	People who work together should get to know each other socially	1 2 3 4 5	I think it is best to keep work and private lives separate
36	I often apologise for things which really were not my fault	1 2 3 4 5	I don't apologise for things which are not my fault
37	I can be rather too stubborn and argumentative	1 2 3 4 5	I am a bit too flexible and adaptable
38	I would rather risk hurting someone than evade the truth	1 2 3 4 5	It is always better to give people support and understanding
39	I do have a tendency to speak out of turn	1 2 3 4 5	I always wait until I am asked to speak
40	I remain the same to deal with even under pressure	1 2 3 4 5	I am easier to get along with once the pressure's off
41	I do not enjoy company social events	1 2 3 4 5	I enjoy meeting colleagues at company social events
42	I am quite easy to convince	1 2 3 4 5	It takes a good argument to make me change my mind
43	I like to do things which benefit other people	1 2 3 4 5	When it comes down to it I'll look after myself first
44	I sometimes feel restless and ill at ease	1 2 3 4 5	I am generally relaxed and carefree
45	I don't mind routine work if I can see a purpose to it	1 2 3 4 5	I have great difficulty in settling down to routine work
46	I find I get inspiration from having people around me	1 2 3 4 5	I like to work alone in order to concentrate
47	I have some very strong views	1 2 3 4 5	I don't have very strong views
48	Worrying is a waste of time	1 2 3 4 5	Worrying about things improves my thinking
49	To be frank some people are beyond help	1 2 3 4 5	All most people need is a little help and understanding

However, the development of a web-based Facet5 questionnaire in 2000 changed the presentation protocol. In the current version items are presented one at a time and people have the opportunity either to answer the question or to skip it and come back to it later.

Web based questions look like this:

Presentation is via an emailed invitation with an embedded link. Clicking on the link brings up an invitation page and further links take the respondent to an introduction/instruction page and then to the questionnaire. The questions are presented one by one and completion is untimed. However response time is recorded.

Initial trials

The first trials of Facet5 occurred between 1988 and 1989. Data was collected from a range of UK businesses and the results of this analysis are provided in detail below.

Development sample

To test the items we presented them to a sample of 693 people from UK organisations. Most of these people were attending assessment centres or employee development centres. The sample was structured as follows:

Development sample by age

Age band (up to)	Number
20	110
25	218
30	74
35	52
40	55
45	50
50	9
55	1
60	3
Unknown	121
Total	**693**

Development sample by gender

Gender	Frequency	Percent
Female	226	32.6%
Male	467	67.4%

Development sample by educational level

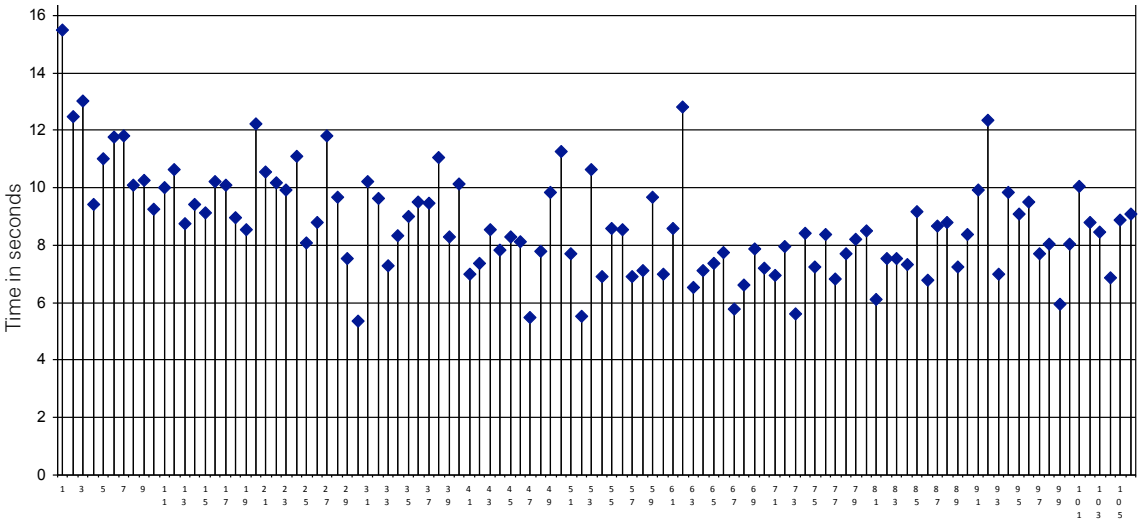
Education	Frequency	Percent
Unknown	51	8.2%
Below O' Levels	9	1.5%
O' Levels	44	7.1%
A' Levels	68	11.0%
Degree	399	64.4%
Higher Degree	49	7.9%

Questionnaire completion time – paper versus web

Experience with many thousands of paper based Facet5 questionnaires indicates that a completion time of about 20–25 minutes was normal. This obviously varied from person to person but in general a completion time of more than 30 minutes was thought to be quite long.

Evidence from web-based completion suggests that the response time is much faster[13]. The chart below indicates the response times for 7564 respondents who completed Facet5 on-line. From this chart it is apparent that most items are responded to in less than 10 seconds (average response latency = 8.8 seconds) giving an overall test time of about 15 minutes. Therefore the combination of sophisticated item format and web-administration leads to a significant decrease in test taking time.

Mean response latency for Facet5 items



n = 7564

Item analysis

Having trialled the 106-item set, descriptive statistical analysis was undertaken to determine a) means, and b) standard deviations. The results are shown below:

Number of items	106
Number of respondents	693
Mean response	2.98
Standard Deviation of responses	0.72

These results helped us to focus attention on the items with the most extreme mean scores. Extremely skewed items suggested that the items were not properly semantically balanced, and as a result one end of the scale was selected disproportionally more often. Specifically, we searched for items where the mean score for the item was within the overall mean score plus or minus 1.5 standard deviations.

Developing Simple Structure

'Simple structure' is the idea that each item relates to only one personality factor. Simple structure is important because it means that the scales themselves are less likely to be correlated. However, it is almost impossible to get pure 'simple structure' where an item loads entirely on one factor and not at all with any other factor. There is always an element of 'noise' in the measurement of any human variable. In our development of the Facet5 questionnaire, we made an absolute correlation of 0.3 as a cut-off point to indicate the acceptable limit. We ensured that each item had a significantly higher correlation than 0.3 on its 'home' factor and a significantly lower correlation with its 'non-home – factors. Any items that did not meet these criteria would lead to domains themselves becoming correlated, and so were excluded from the scoring keys.

Item 28 is one example. It correlates – 0.47 with Will but also 0.32 with Affection. Item 16 correlates 0.32 with Emotionality and – 0.31 with Energy. Items like this would lead to domains becoming correlated. In both cases these items were excluded from the scoring keys.

However the final decision on item inclusion was not made on a strictly actuarial basis. Some items that may have been skewed were retained due to their clean loading on their home factor.

The result of this was that 23 items were screened out, leaving us with 83 items that were assigned to the personality factors in the following way:

Facet5 Factor	Number of items
Will	15
Energy	15
Affection	18
Control	17
Emotionality	18

The result of this pruning exercise is that Facet5 has simple structure. Of the 83 items included, none have absolute correlations greater than 0.3 with a factor other than their 'home' factor. The final scoring keys include all of these 83 items. The scoring algorithms produce a single raw score for each factor. The following table shows the item analysis from 6534 cases:

Item	Will	Energy	Affection	Control	Emotionality
1			-0.42		
2				-0.33	
3				0.46	
4					-0.42
5	-0.51				
6			0.54		
7				0.44	
8	-0.06	0.11	0.07	0.15	-0.12
9		0.44			
10	0.14	-0.03	-0.20	-0.08	0.01
11			-0.31		
12	-0.02	0.14	0.10	0.23	-0.15
13				-0.54	
14					0.30
15	-0.23	0.12	0.12	0.04	0.01
16					0.32
17			0.33		
18				0.40	
19		0.32			
20					-0.40
21	0.24	0.06	-0.06	0.10	-0.10
22			-0.47		
23				-0.48	
24	-0.09	-0.16	0.03	0.17	0.11
25	-0.35				
26	-0.47				
27	-0.09	0.10	0.20	0.25	-0.17
28	0.20	0.21	-0.11	-0.24	-0.20
29				0.44	

Item	Will	Energy	Affection	Control	Emotionality
30		0.61			
31			-0.33		
32					0.58
33			-0.40		
34				-0.41	
35		-0.35			
36	0.23	0.06	-0.16	0.00	-0.22
37	-0.56				
38	-0.47				
39	-0.40				
40					0.52
41		0.46			
42	0.42				
43			-0.56		
44					-0.55
45				-0.46	
46		-0.47			
47	-0.44				
48					0.38
49			0.55		
50				0.45	
51					-0.50
52					-0.61
53					
54			-0.43		
55				-0.31	
56					0.50
57		-0.56			
58	-0.45				
59	-0.18	-0.16	0.09	0.08	0.19
60				0.38	
61				0.42	

Item	Will	Energy	Affection	Control	Emotionality
62		0.48			
63			-0.30		
64					0.39
65			-0.56		
66	-0.15	0.03	0.27	-0.09	-0.06
67		-0.47			
68					-0.34
69	-0.37				
70			0.31		
71				0.41	
72					0.39
73		0.53			
74	0.48				
75			-0.36		
76	0.01	-0.05	-0.07	-0.24	-0.10
77				-0.44	
78		-0.37			
79	-0.45				
80					0.31
81			0.44		
82				0.43	
83		0.48			
84					-0.38
85					-0.38
86	0.23	0.13	-0.17	0.06	-0.22
87				-0.34	
88	0.06	0.10	-0.08	0.00	0.02
89		-0.34			
90	-0.54				
91			0.56		
92					-0.48
93			-0.45		

Item	Will	Energy	Affection	Control	Emotionality
94			-0.32		
95	0.36				
96	-0.12	-0.11	0.14	0.25	0.07
97			-0.40		
98				-0.47	
99					0.34
100	-0.04	0.20	0.12	-0.02	-0.23
101	-0.31				
102			0.51		
103				0.41	
104					0.41
105		0.32		-0.40	
106					

The items with the heaviest loading are highlighted. The 23 items that are not included in the existing scoring algorithm are retained as placeholders and are shaded in this table. As part of Facet5's continuing research program, these items are swapped out from time to time and replaced with research items. For example we recently inserted the items from Judge's Core Self Evaluation Scale to see how it related to Facet5[14].

This analysis gave us confidence that the selected items would provide a sound basis for the measurement of the 5 core factors. Our next step was to evaluate each factor-based item set (eg the items that related to Will) in more detail to see whether they were measuring a single core construct (they were uni-dimensional) or whether in fact there was a more complex picture embedded (they were multi-dimensional).

Scoring main Factors

The Facet5 main factor raw scores are computed simply by adding up the scores given to each item. However many items need to be reversed so the process is as follows:

Step 1: extract the items which load on the factor

Step 2: reverse the items which are reversed scored. This is done by subtracting the response given from 6. So if a respondent has answered 5 on a scale that need to be reversed, then this score becomes 1 (6 – 5 = 1).

Step 3: sum the items to get a raw score for that factor

Step 4: raw scores are then standardised using values from the selected norm group (see later)

Distribution of Factor scores

A recent analysis of a much larger sample (n = 33333) shows the distribution of the scores on the five main Facet5 factors. This sample was selected from a UK data set and all respondents completed the questionnaire in English. Note that the introduction to the questionnaire allows the respondent to choose an alternative language and therefore the fact that English was chosen when alternatives were available suggests that for this sample English was either the mother tongue or the language in which they felt most comfortable.

Descriptive statistics for main Factors

For each factor the following results show the summary statistics, the frequency distribution of the raw scores and a histogram to test normality and to provide a visual indication of the distribution.

Each of these analyses shows that the scores on the five main dimensions assessed by Facet5 will spread along the scale according to the normal distribution curve, or Gaussian distribution. Most people will have scores in the middle and progressively fewer will have extreme scores. The Facet5 Questionnaire is normative, in that results are expressed by comparison to a particular reference group or 'norm.' The scores obtained allow comparison from one individual to another. They provide a statement of how each individual scores compared to their reference group on each of the factors. The distributions for the raw scores for the evaluation sample (UK-based, English Speaking) are shown below:

Will – Descriptive statistics

N	Valid	33330
Mean		46.14
Standard Error of Mean		.039
Standard Deviation		7.155
Skewness		.082
Standard Error of Skewness		.013
Kurtosis		.167
Standard Error of Kurtosis		.027
Minimum		19
Maximum		75

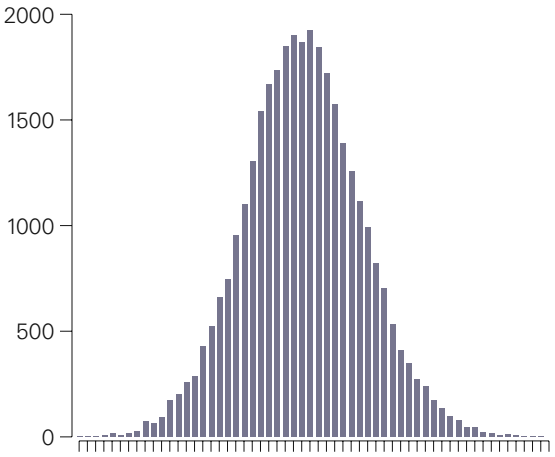
Will – Frequency distribution

Raw score	Frequency	Percent	Cumulative percent
19	3	0.01	0.01
20	1	0.00	0.01
21	4	0.01	0.02
22	7	0.02	0.05
23	17	0.05	0.10
24	8	0.02	0.12
25	18	0.05	0.17
26	25	0.08	0.25
27	73	0.22	0.47
28	65	0.20	0.66
29	93	0.28	0.94
30	175	0.53	1.47
31	200	0.60	2.07
32	258	0.77	2.84
33	285	0.86	3.70
34	430	1.29	4.99
35	525	1.58	6.56
36	660	1.98	8.54
37	747	2.24	10.78
38	955	2.87	13.65
39	1102	3.31	16.95
40	1305	3.92	20.87
41	1544	4.63	25.50
42	1670	5.01	30.51
43	1735	5.21	35.72
44	1850	5.55	41.27
45	1903	5.71	46.98
46	1869	5.61	52.59
47	1924	5.77	58.36
48	1846	5.5	63.9
49	1723	5.2	69.1
50	1577	4.7	73.8
51	1389	4.2	78.0
52	1256	3.8	81.7
53	1114	3.3	85.1
54	992	3.0	88.1

Raw score	Frequency	Percent	Cumulative percent
55	821	2.5	90.5
56	704	2.1	92.6
57	534	1.6	94.2
58	412	1.2	95.5
59	350	1.1	96.5
60	273	.8	97.3
61	238	.7	98.0
62	175	.5	98.6
63	136	.4	99.0
64	95	.3	99.3
65	78	.2	99.5

Raw score	Frequency	Percent	Cumulative percent
66	47	.1	99.6
67	45	.1	99.8
68	23	.1	99.8
69	17	.1	99.9
70	7	.0	99.9
71	12	.0	100.0
72	7	.0	100.0
73	3	.0	100.0
74	3	.0	100.0
75	2	.0	100.0
Total	33330	100.0	

Will – Histogram of raw scores



Will – Summary

The distribution of Will is clearly normal although it is slightly leptokurtic (Kurtosis = 0.167) but is not skewed to any noticeable degree (Skewness = 0.082). With a mean score of 46.14 and an SD of 7.15, 68% of scores will lie between 40 and 53. Scores outside of this range would be classed as unusual (either high or low).

Energy – Descriptive statistics

N	Valid	33330
Mean		51.87
Standard Error of Mean		.044
Standard Deviation		7.969
Skewness		-.262
Standard Error of Skewness		.013
Kurtosis		-.020
Standard Error of Kurtosis		.027
Minimum		19
Maximum		75

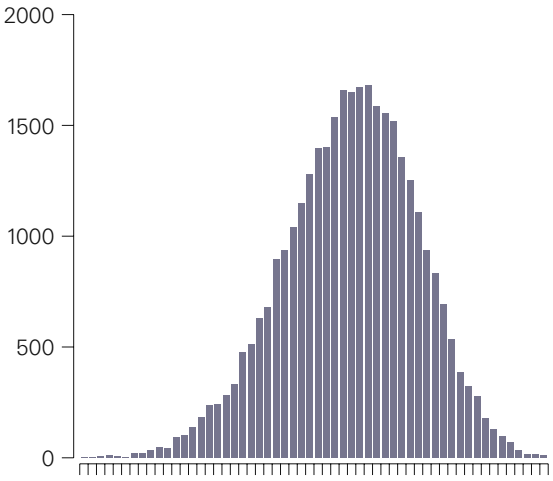
Energy – Frequency distribution

Raw score	Frequency	Percent	Cumulative percent
19	1	0.00	0.00
21	1	0.00	0.01
22	7	0.02	0.03
23	11	0.03	0.06
24	5	0.02	0.08
25	4	0.01	0.09
26	19	0.06	0.14
27	21	0.06	0.21
28	33	0.10	0.31
29	47	0.14	0.45
30	43	0.13	0.58
31	94	0.28	0.86
32	102	0.31	1.16
33	138	0.41	1.58
34	184	0.55	2.13
35	239	0.72	2.85
36	242	0.73	3.57
37	284	0.85	4.43
38	332	1.00	5.42
39	476	1.43	6.85
40	511	1.53	8.38
41	629	1.89	10.27
42	679	2.04	12.31
43	898	2.69	15.00
44	939	2.82	17.82
45	1039	3.12	20.94
46	1149	3.45	24.38
47	1280	3.84	28.22
48	1398	4.19	32.42
49	1402	4.21	36.62
50	1538	4.61	41.24
51	1658	4.97	46.21
52	1650	4.95	51.16
53	1673	5.02	56.18
54	1683	5.05	61.23
55	1587	4.76	65.99

Raw score	Frequency	Percent	Cumulative percent
56	1557	4.67	70.67
57	1518	4.55	75.22
58	1357	4.07	79.29
59	1255	3.77	83.06
60	1107	3.32	86.38
61	938	2.81	89.19
62	832	2.50	91.69
63	693	2.08	93.77
64	534	1.60	95.37
65	388	1.16	96.53
66	323	0.97	97.50

Raw score	Frequency	Percent	Cumulative percent
67	279	0.84	98.34
68	178	0.53	98.87
69	129	0.39	99.26
70	99	0.30	99.56
71	70	0.21	99.77
72	34	0.10	99.87
73	18	0.05	99.92
74	14	0.04	99.97
75	11	0.03	100.00
Total	33330	100	

Energy – Histogram of raw scores



Energy – Summary

The distribution of Energy is nearly normal in terms of Kurtosis (Kurtosis = -.020) but is slightly negatively skewed (Skewness = -0.262). With a mean score of 51.87 and an SD of 7.97, 68% of scores should lie between 43.9 and 59.8. Examination of the frequency distribution suggests that in fact in spite of the slight negative skew, the cut-off point for the lowest 16% of cases is somewhere between 43 and 44. Therefore there is little difference between the estimate that would be made by assuming normality and by a more detailed scaling using the cumulative frequency distribution.

Affection – Descriptive statistics

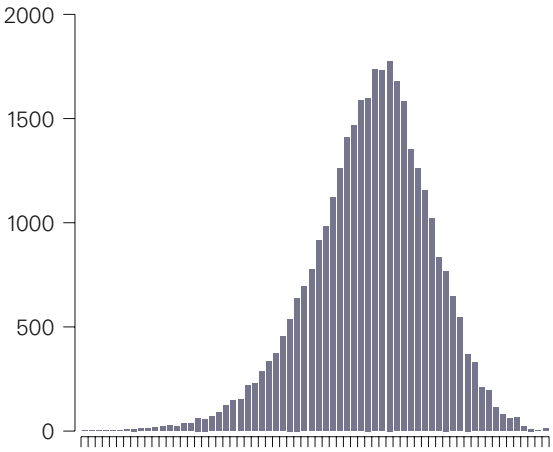
N	Valid	33330
Mean		65.38
Standard Error of Mean		.046
Standard Deviation		8.390
Skewness		-.460
Standard Error of Skewness		.013
Kurtosis		.562
Standard Error of Kurtosis		.027
Minimum		20
Maximum		90

Affection – Frequency distribution

Raw score	Frequency	Percent	Cumulative percent
20	1	0.00	0.00
23	1	0.00	0.01
26	1	0.00	0.01
27	2	0.01	0.02
28	4	0.01	0.03
30	3	0.01	0.04
31	6	0.02	0.05
32	10	0.03	0.08
33	12	0.04	0.12
34	13	0.04	0.16
35	17	0.05	0.21
36	21	0.06	0.27
37	26	0.08	0.35
38	23	0.07	0.42
39	35	0.11	0.53
40	37	0.11	0.64
41	63	0.19	0.83
42	58	0.17	1.00
43	69	0.21	1.21
44	90	0.27	1.48
45	124	0.37	1.85
46	145	0.44	2.28
47	151	0.45	2.74
48	220	0.66	3.40
49	229	0.69	4.08
50	285	0.86	4.94
51	333	1.00	5.94
52	373	1.12	7.06
53	454	1.36	8.42
54	538	1.61	10.03
55	639	1.92	11.95
56	692	2.08	14.03
57	775.00	2.33	16.35
58	914	2.74	19.09
59	980	2.94	22.03
60	1121	3.36	25.40
61	1259	3.78	29.17
62	1409	4.23	33.40

Raw score	Frequency	Percent	Cumulative percent
63	1465	4.40	37.80
64	1587	4.76	42.56
65	1599	4.80	47.36
66	1736	5.21	52.57
67	1730	5.19	57.76
68	1776	5.33	63.08
69	1676	5.03	68.11
70	1580	4.74	72.85
71	1352	4.06	76.91
72	1261	3.78	80.69
73	1155	3.47	84.16
74	1019	3.06	87.22
75	833	2.50	89.71
76	768	2.30	92.02
77	647	1.94	93.96
78	545	1.64	95.60
79	370	1.11	96.71
80	328	0.98	97.69
81	210	0.63	98.32
82	193	0.58	98.90
83	113	0.34	99.24
84	78	0.23	99.47
85	59	0.18	99.65
86	67	0.20	99.85
87	23	0.07	99.92
88	10	0.03	99.95
89	4	0.01	99.96
90	13	0.04	100.00
Total	33330	100.00	

Affection – Histogram of raw scores



Affection – Summary

Affection appears to be both leptokurtic and negatively skewed. This would suggest that there are slightly more people with high Affection (the positive shoulder of the distribution is relatively high) but there are still a number of people with very low Affection scores. The assumption of normality would suggest that 16% of the respondents will have Affection raw scores below 56.99 (Mean(65.38) –1 Standard Deviation (8.39)) whereas scaling using the cumulative frequency distribution would suggest dropping this by perhaps 1 raw score 56. Therefore in spite of the negative skew there would seem to be little to be gained by a different method of scaling.

Control – Descriptive statistics

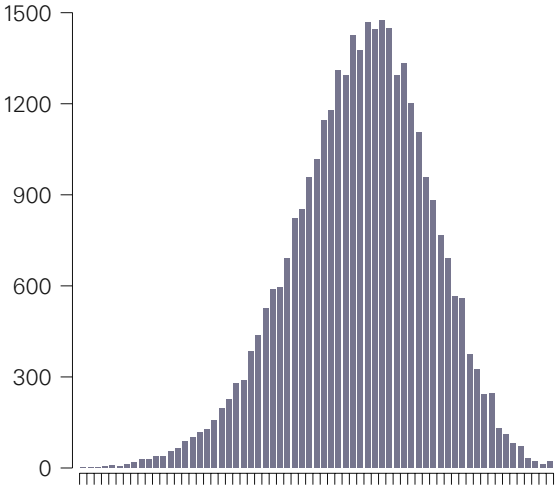
N	Valid	33330
Mean		59.06
Standard Error of Mean		.051
Standard Deviation		9.353
Skewness		-.320
Standard Error of Skewness		.013
Kurtosis		.083
Standard Error of Kurtosis		.027
Minimum		21
Maximum		85

Control – Frequency distribution

Raw score	Frequency	Percent	Cumulative percent
21	1	0.00	0.00
22	1	0.00	0.01
23	1	0.00	0.01
24	5	0.02	0.02
25	9	0.03	0.05
26	6	0.02	0.07
27	10	0.03	0.10
28	17	0.05	0.15
29	28	0.08	0.23
30	27	0.08	0.32
31	37	0.11	0.43
32	39	0.12	0.54
33	55	0.17	0.71
34	65	0.20	0.90
35	86	0.26	1.16
36	99	0.30	1.46
37	117	0.35	1.81
38	128	0.38	2.19
39	158	0.47	2.67
40	196	0.59	3.26
41	226	0.68	3.93
42	279	0.84	4.77
43	287	0.86	5.63
44	385	1.16	6.79
45	436	1.31	8.09
46	526	1.58	9.67
47	589	1.77	11.44
48	595	1.79	13.23
49	689	2.07	15.29
50	823	2.47	17.76
51	850	2.55	20.31
52	957	2.87	23.18
53	1018	3.05	26.24
54	1145	3.44	29.67
55	1177	3.53	33.20
56	1310	3.93	37.13
57	1293	3.88	41.01
58	1426	4.28	45.29

Raw score	Frequency	Percent	Cumulative percent
59	1377	4.13	49.42
60	1468	4.40	53.83
61	1445	4.34	58.16
62	1475	4.43	62.59
63	1449	4.35	66.94
64	1293	3.88	70.82
65	1334	4.00	74.82
66	1200	3.60	78.42
67	1105	3.32	81.73
68	958	2.87	84.61
69	881	2.64	87.25
70	767	2.30	89.55
71	691	2.07	91.63
72	566	1.70	93.32
73	559	1.68	95.00
74	375	1.13	96.13
75	326	0.98	97.10
76	241	0.72	97.83
77	247	0.74	98.57
78	131	0.39	98.96
79	110	0.33	99.29
80	80	0.24	99.53
81	72	0.22	99.75
82	30	0.09	99.84
83	20	0.06	99.90
84	13	0.04	99.94
85	21	0.06	100.00
N = 33333			

Control – Histogram of raw scores



Control – Summary

The distribution for Control ranges from 21 to 85 with a mean of 59.06. The SD is 9.35. It is slightly negatively skewed (-0.32) and is slightly leptokurtic (0.083). As with the other factors, algebraic scaling using the normal distribution would suggest that 16% of the responses would have raw scores below 49.7. Cumulative frequency scaling would provide almost exactly the same result.

Emotionality – Descriptive statistics

N	Valid	33330
Mean		47.41
Standard Error of Mean		.053
Standard Deviation		9.720
Skewness		.246
Standard Error of Skewness		.013
Kurtosis		-.044
Standard Error of Kurtosis		.027
Minimum		18
Maximum		85

Emotionality – Frequency distribution

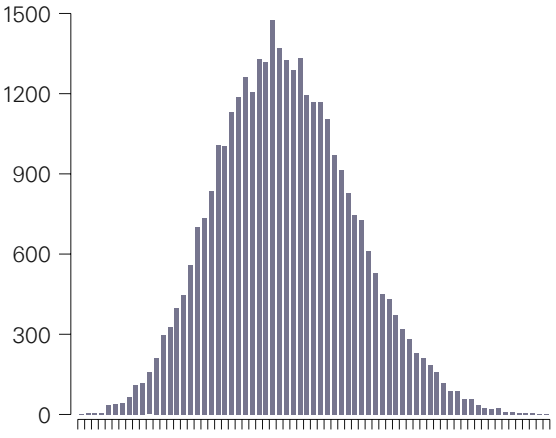
Raw score	Frequency	Percent	Cumulative percent
18	3	0.01	0.01
19	7	0.02	0.03
20	4	0.01	0.04
21	5	0.02	0.06
22	36	0.11	0.17
23	39	0.12	0.28
24	44	0.13	0.41
25	65	0.20	0.61
26	109	0.33	0.94
27	117	0.35	1.29
28	157	0.47	1.76
29	213	0.64	2.40
30	298	0.89	3.29
31	328	0.98	4.28
32	398	1.19	5.47
33	447	1.34	6.81
34	560	1.68	8.49
35	702	2.11	10.60
36	735	2.21	12.80

Raw score	Frequency	Percent	Cumulative percent
37	837	2.51	15.31
38	1007	3.02	18.33
39	1003	3.01	21.34
40	1130	3.39	24.73
41	1188	3.56	28.30
42	1263	3.79	32.09
43	1206	3.62	35.71
44	1331	3.99	39.70
45	1318	3.95	43.65
46	1477	4.43	48.09
47	1370	4.11	52.20
48	1326	3.98	56.17
49	1287	3.86	60.04
50	1333	4.00	64.04
51	1196	3.59	67.62
52	1168	3.50	71.13
52	1168	3.50	71.13
53	1107	3.32	74.45
54	970	2.91	77.36

Raw score	Frequency	Percent	Cumulative percent
55	915	2.75	80.11
56	827	2.48	82.59
57	745	2.24	84.82
58	729	2.19	87.01
59	612	1.84	88.84
60	529	1.59	90.43
61	450	1.35	91.78
62	431	1.29	93.08
63	371	1.11	94.19
64	321	0.96	95.15
65	281	0.84	95.99
66	230	0.69	96.68
67	213	0.64	97.32
68	184	0.55	97.88
69	160	0.48	98.36
70	117	0.35	98.71

Raw score	Frequency	Percent	Cumulative percent
71	88	0.26	98.97
72	87	0.26	99.23
73	59	0.18	99.41
74	57	0.17	99.58
75	34	0.10	99.68
76	25	0.08	99.76
77	21	0.06	99.82
78	23	0.07	99.89
79	10	0.03	99.92
80	8	0.02	99.94
81	6	0.02	99.96
82	4	0.01	99.97
83	6	0.02	99.99
84	2	0.01	100.00
85	1	0.00	100.00
Total	33330	100.00	0.00

Emotionality – Histogram of raw scores



Emotionality – Summary

The distribution for Emotionality ranges from 18 to 85 with a mean of 47.41. The SD is 9.72. It is slightly positively skewed (0.246) and has little kurtosis (-0.04). Algebraic scaling (mean + 1SD) would suggest that 16% of the responses would have raw scores above 57.

The cumulative frequency list shows that 15.2% of the sample have raw scores above 57 so the two methods would give almost exactly the same result. There would appear to be little difference between the two methods.

Facet5 sub-factors

When we define a domain there is always a trade off. We want a set of items that reflect the core of the factor but we also want sufficient range to capture the breadth. This is considered carefully during domain definition and item selection. It is sometimes referred to as the 'fidelity – bandwidth' trade off. It emerged in the early development of Facet5 where some people suggested that a more 'fine grained' tool which had more factors would provide greater understanding.

Each of the core personality factors contains a discrete number of sub-factors. Will for example, covers behaviours that include Determination, Confrontation and Independence. Developing items that fit particular personality factors is a careful process, as items must capture the core of the factor as well as its sub-factors. When completing a Facet5 personality questionnaire, a person might select the items relating to Confrontation more heavily than those relating to Independence. In this instance, it would be helpful to delineate the sub-factors during feedback.

Sub-factors can be created in at least three different ways:

- By creating a full set of items that are specifically keyed to a predefined sub-factor. This approach can be seen in the NEO-PI[15]. For example, the main factor of Extraversion is made up of six sub-factors and each of these sub-factors has 8 items.
- By creating a small set of items that are very close in meaning and which represent a particular construct of interest. This is the approach taken, for example, by the Comrey Personality Scales[16] and by the Hogan Personality Inventory (HPI)[17]. Comrey refers to these sets of items as Factored Homogeneous Item Dimensions (FHID) while Hogan calls them Homogeneous Item Clusters or 'HICs.' Some HICs may have only 3 or 4 items.
- By completing a second-level analysis (factor analysis) of the items already identified as belonging to a single broad construct. This is the approach taken by Facet5. For example, during Facet5's development the 15 items that relate to Will were subjected to a Principal Components analysis to see whether they could be divided into groups that were conceptually similar. We follow this with a Promax rotation to make the factors more easily interpretable. The process for this analysis is described more fully elsewhere[18].

This third approach provides a result that is somewhere between the previous two. A Promax rotation is defined as providing 'an oblique approximation to simple structure.' (As aforementioned, simple structure refers to the situation where an item loads on one scale only – for example if an item is thought to measure Will then it scores on Will and no other factor.) The Promax method provides an estimate of this. It produces a set of factor loadings where the most important items for a particular factor have a high loading and those that are less important have a comparatively low loading. The weights therefore are not 1 and 0 (which would be perfect simple structure) but rather are close to 1 and close to 0. As a result, Promax rotated factors are slightly correlated, but this ultimately makes them easier to understand and interpret. Essentially, the Promax method is thought to be a better reflection of the real world. Hans Eysenck once said that 'the whole world correlates about 0.3'[19].

Sample size required for analysis

There is debate about what sample size is required to conduct a reliable factor analysis. Osborne and Costello[20] conducted a study on appropriate sample size; their investigative questions being:

- Is it just a matter of having a particular sample size? or;
- Is it more to do with the ratio between the number of variables and the number of cases?

In this study Osborne and Costello pointed out that it was generally accepted in the psychological community that a sample of 100 is 'poor' while a sample size of 1000 is 'excellent'. Osborne and Costello further concluded that there was a general agreement that in order for data analysis to be valid, an experimenter needed approximately 15–30 times as many cases as he or she had variables. In the end they came to the conclusion that, when analysing any type of data, 'more is always better.'

Applying Orsborne and Costello's findings to the development of Facet5, we realised that if Facet5 has 106 variables, we would need at least 3000 cases of data for our analysis to be valid. The main Facet5 research database is dynamic and constantly growing so for this analysis we were able to extract a large database of 33,333 cases of people who had completed the questionnaire in English and who were from the UK. Using cases from one country only was intended to reduce any possible cultural effects. Within this sample, 49% were male, 38% female and the remainder were of unknown gender. We analysed each of the main Facet5 factors in turn to determine:

- How the main factors were distributed
- The dimensionality of each of the main factors
- The relationship between sub-factors and main factors

The results were as follows:

Will sub-factors

Will is measured with 15 items. The first analysis suggested that there were 4 factors with Eigen values greater than 1. However the last of these values was 1.04 and proved impossible to interpret so we retained only the first three sub-factors for use. They provided a useful insight into the structure of the Will items. The table below shows the distribution of the highest loading items across the three sub-factors Determination, Confrontation and Independence:

Will – showing sample items	Sub-factor loading		
Item	Determination	Confrontation	Independence
I have been accused of being stubborn and inflexible	.640		
It takes a good argument to make me change my mind	.602		
I will not sacrifice my principles	.577		
I believe in telling people exactly how I feel		.683	
I have often been the first to take responsibility for things		.601	
I am often among the first to speak out on a topic		.515	
I am particularly good at arguing a case aggressively			.639
If a problem arises with someone at work I act fast and hard			.616
It is important for a manager to set very clear goals for others			.474

Full interpretation of these sub-factors can be found in other sections of this manual.

Energy sub-factors

Energy is also measured with 15 items. There were 4 factors with Eigen values greater than 1 however the last of these was 1.006 and proved impossible to interpret. The remaining three sub-factors were retained for Energy. The table below shows the distribution of the highest loading items across the three sub-factors Vitality, Sociability, Adaptability:

Energy – showing sample items	Sub-factor loading		
Item	Vitality	Sociability	Adaptability
It is the risk that makes the job interesting	0.60		
I am at my best when defending my ideas on my feet	0.57		
I enjoy being the centre of attention	0.55	0.49	
I enjoy meeting colleagues at company social events		0.76	
I enjoy social activities		0.75	
I make new friends easily		0.70	
I find I get inspiration from having people around me		0.39	0.62
I like to have someone I can bounce ideas off			0.73
I prefer to adapt my ideas to changing circumstances			0.54

Affection sub-factors

Affection is measured with 18 items. There were 5 factors that had Eigen values greater than 1 but the 4th and 5th factors were extremely complex with items being spread across more than one sub-factor. This made them very difficult to interpret. When we constrained the result to two and three factors the items were grouped in a much more meaningful way. We therefore decided to retain a three factor solution for Affection. This structure is shown in the following table that shows the heaviest loading items on each of the three derived factors Altruism, Support and Trust:

Affection – showing sample items	Sub-factor loading		
Item	**Altruism**	**Support**	**Trust**
I put other people's needs ahead of my own	0.79		
I like to do things which benefit other people	0.69		
We should consider other people's needs first	0.68		
I get great satisfaction out of being able to help other people	0.57		
I don't like people who take advantage of a situation	0.53		
All most people need is a little help and understanding		0.66	
Everyone has their good points		0.62	
I don't like to judge other people		0.61	
Everybody has the right to a decent job		0.55	
Most people can be trusted			0.64
In business most people are prepared to help others			0.52
Social welfare programmes are a sign of a civilised society			0.50

Control sub-factors

Control is measured with 17 items. There were 4 factors that had Eigen values greater than 1 but the scree plot showed a marked kink after the second factor. Given that the 3rd and 4th factors were obscure and complex we retained a 2-factor solution. The factors with the highest loading items were Discipline and Responsibility and are shown in the following table:

Control – showing sample items	Sub-factor loading	
Item	**Discipline**	**Responsibility**
I enjoy following a task through to the finish	0.73	
I can't leave a task until it is finished	0.72	
I don't mind routine work if I can see a purpose to it	0.58	
I insist on having a plan before acting	0.51	
There are certain codes of conduct which are absolute		0.63
It is important to have rules so people know where they stand		0.55
Appropriate dress for work is important		0.53

Emotionality sub-factors

The original research identified 18 items that loaded on Emotionality. There were 5 factors that had Eigen values greater than 1, but the last three were very spread out and proved difficult to interpret. However it is possible that further research and the trialling of new items could expand our interpretation of Emotionality. In the meantime the two factors that emerged clearly were Tension and Apprehension and are shown below:

Emotionality – showing sample items	Sub-factor loading	
Item	**Tension**	**Apprehension**
Worrying about things improves my thinking	0.67	
As an important event gets closer I tend to worry more about it	0.61	
I don't know how some people can relax despite any problems	0.60	
Having done my best I still can't relax until my review is over	0.52	
I would call myself a nervous person	0.52	
I sometimes feel restless and ill at ease	0.50	
I am easier to get along with once the pressure's off		0.59
I prefer to have peace and quiet if I have to concentrate		0.55
I hate being interrupted in my work		0.55

Summary of Facet5 sub-factors

The net result of these analyses was the determination of the following Factors and sub-factors for Facet5:

Factor	Sub-factor	Description
Will	Determination	The inner drive to commit to own ideas
	Confrontation	A drive to confront issues as they arise
	Independence	A tendency to go your own way
Energy	Vitality	Obvious enthusiasm and energy
	Sociability	Interest in being with people
	Adaptability	Involving others in your thinking
Affection	Altruism	Putting other people's interests first
	Support	Always trying to be understanding
	Trust	Tendency to take people at face value
Control	Discipline	Being personally organised and planned
	Responsibility	Being willing to take personal responsibility
Emotionality	Anxiety	A general sense of tension or stress
	Apprehension	Being cautious and not over-optimistic

Scoring Facet5 sub-factors

Although the tables above show the key items loading on each sub-factor, sub-factor scores are based on the factor estimation weights for all of the items in the scale, which in turn are linked to the factor loadings determined by the original analysis. As a result each sub-factor score is a differentially weighted linear sum of all the items within that factor. For example Determination is:

Determination = (ZQ47*0.29096)+(ZQ26*0.22117)+(ZQ21*0.26668)+(ZQ95*0.29670)+(ZQ42*0.31770)+(ZQ79*0.15677)+(ZQ101*0.23596)+(ZQ37*0.03712)+(ZQ90R*0.15127)+(ZQ25*0.07886)+(ZQ74*0.03632)+(ZQ15*0.11445)+(ZQ38*0.06359)+(ZQ58*0.09767)+(ZQ5*0.10011)

Note that the variables in this equation are prefixed with 'Z'. This indicates that they are standardised scores computed from a specific norm group. For example, in order to convert a value for Item 47, we would need to know the mean and standard deviation for all responses to Item 47 in the relevant norm group. Once we have computed the values for each sub-factor they are then standardised as described later.

Distribution of sub-factor scores

Since the sub-factor scores are computed from the differentially weighted sum of the standardised item scores, it would be expected that they would be distributed normally in the same way as the main factor scores. Obviously they may be moved up or down the scale due to different mean scores and standard deviations.

Descriptive statistics for sub-factors

For each sub-factor the following results show the summary statistics for the sub-factor scores prior to standardisation and conversion to Sten scores.

Facet 5 sub-factors – raw scores

N = 33333	Mean	Standard Deviation	Minimum	Maximum
W1	-0.03	0.94	-2.81	4.82
W2	0.15	0.89	-4.06	3.13
W3	0.07	0.94	-3.58	2.97
E1	-0.01	1.00	-3.84	3.25
E2	-0.01	0.99	-3.93	2.48
E3	-0.06	0.91	-2.87	4.78
A1	-0.07	0.92	-3.53	4.48
A2	0.10	0.75	-3.12	4.88
A3	-0.07	0.94	-3.23	5.30
C1	-0.09	1.00	-4.05	2.93
C2	0.05	0.97	-3.29	4.91
Em1	-0.01	0.95	-3.56	2.77
Em2	-0.06	0.97	-4.86	3.16

Facet5 sub-factor labels				
W1 Determination	E1 Vitality	A1 Altruism	C1 Discipline	Em1 Tension
W2 Confrontation	E2 Sociability	A2 Support	C2 Responsibility	Em2 Apprehension
W3 Independence	E3 Adaptability	A3 Trust		

Note that these are computed from the standardised raw scores and therefore we would expect the means and SDs to be close to 0 and 1 respectively.

Standardisation of Facet5 raw scores

Reporting raw scores can be a problem for a number of reasons. Telling someone their Will score was 45 is of little value unless they know how that compares to others. There is also a problem when you look at raw scores across a number of factors. Each factor may have a different score distribution so that a score of 45 may be high for one factor but average for another.

One solution is to standardise the raw scores. Standardisation involves computing the average score and the standard deviation for a group, and then using this information to convert all the factors to a common baseline[21]. Common standardisation methods include:

z-scores

A z-score converts the raw score to have a mean of zero and a standard deviation of 1. Scores then range above and below zero. Although technically perfectly valid, z-scores can seem awkward to work with because of the fact that they can be small negative decimals such as -1.3. As a result most tools convert the data into more manageable numbers. The process simply involves computing the mean and standard deviation for the norm group and then applying the following formula:

$$z = \frac{Rawscore - Mean}{Standard\ deviation}$$

There are three elements to this formula that are critical:

1 The respondent's raw score which is based on the actual responses to the questionnaire

2 The average score for the comparison group

3 The standard deviation of the comparison group

From this there three common transformations that are made to make the scores easier to understand.

Stanine scores (Standard Nine)

A stanine standardisation starts with a z-score and then transposes it so it has a mean of 5 and a standard deviation of 2. Using this method scores range between 1 and 9. Although it is technically possible to score below 1 or above 9, such extreme scores are usually truncated to 1 or 9.

t-scores

A t-score is a transformation which gives score which have a mean of 50 and a standard deviation of 10. Scores then range from 1 to 100. Scores beyond this range are truncated.

Sten Scores (Standard Ten)

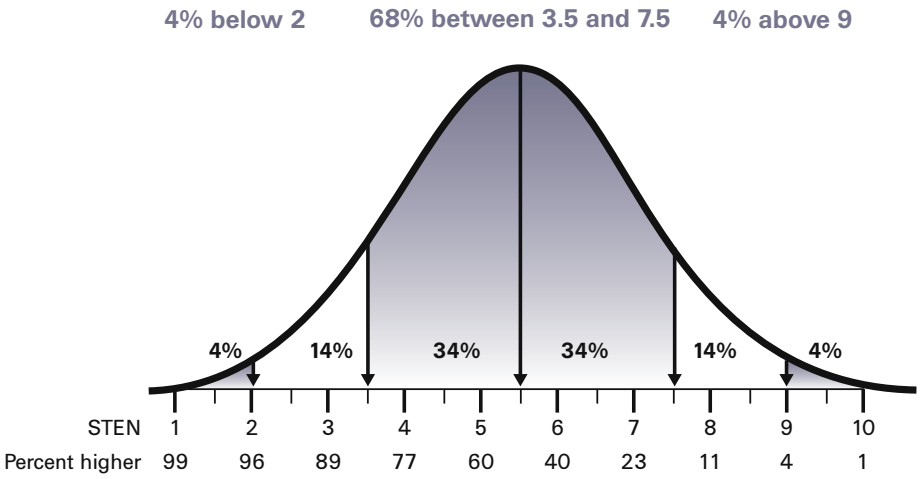
Sten scores also start with a z-score but then they convert it to a mean of 5.5 and a standard deviation of 2. Scores are truncated to between 1 and 10. This is the standardisation used by Facet5. After the raw score is computed the system applies the following formula:

$$\text{Sten score} = \left(\left(\left(\frac{\text{Raw score} - \text{Average score}}{\text{Standard deviation}}\right) * 2\right) + 5.5\right)$$

The scoring program automatically applies this formula and converts the scores to 'Sten scores'.

Distribution of Sten scores

When a sample is tested and their sten scores are charted they will form a 'normal distribution' such as the one shown below:



From this you can see that a score between 3.5 and 7.5 is obtained by about 68% (about two thirds) of the sample. You might call that an average score. But a score that is 2 SDs above the mean (ie 9.5), is only obtained by 2.5% of the population so is rare or unusual.
The same can be said of scores below 1.5.

As individual scores reach these extremes, so the typical behaviour more obviously reflects the core elements of the factor.

Relationship between Facet5 factors

For a personality questionnaire to be effective, it is not only important that it has sufficient factors to account for most of the variation between people, but also that the factors that it produces are not heavily overlapping. There is little point in measuring the same thing a number of times over. The degree to which a questionnaire does this is demonstrated by the correlations between the factors. In general terms, a correlation of less than 0.30 would be considered low enough to suggest that the factors are not duplicated. The correlations between Facet5 factors for the UK English research sample are shown below:

N = 33333	Facet5 – inter correlations			
	Energy	Affection	Control	Emotionality
Will	0.31	-0.33	-0.09	-0.20
Energy		0.01	-0.14	-0.38
Affection			0.20	-0.07
Control				0.04
Average absolute correlation: 0.07				

A previous analysis from the development sample gave the following intercorrelations:

N = 693	Facet5 – inter correlations			
	Energy	Affection	Control	Emotionality
Will	0.25	-0.25	0.09	-0.18
Energy		0.03	0.01	-0.30
Affection			0.11	-0.10
Control				0.01
Average absolute correlation: 0.03				

Analyses of other data sets show very similar results. From this it can be seen that the Facet5 factors are largely independent, although there are small negative correlations between Will and Affection and between Energy and Emotionality.

Relationship between sub-factors and main factors

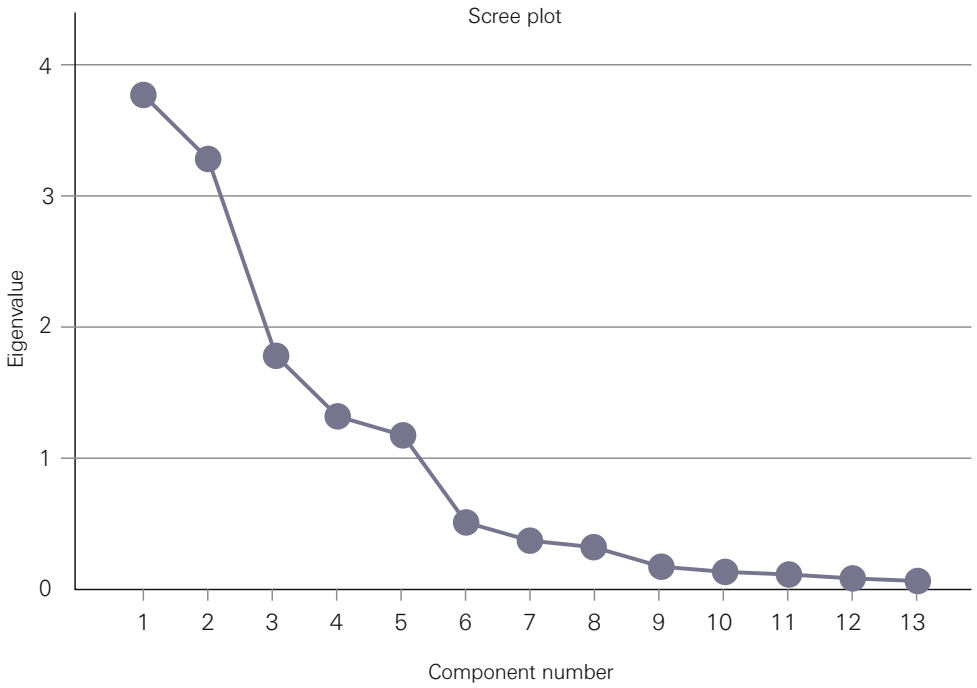
The initial item analysis shows that Facet5 has simple structure: each item loads on only one factor. However, it is also important to have simple structure at the sub-factor level: the sub-factors should load significantly on their home factor but very little on their non-home factors.

We tested whether we had simple structure at the sub-factor level by computing scores for each sub-factor, and subsequently analysing the correlations between these scores. To test the number of factors in this data we used the Kaiser Criterion (sometimes called Little Jiffy), which limits the number of Factors to those that have an Eigen value greater than 1. We also applied Cattell's scree test by plotting the Eigen values on a chart and looking for a noticeable kink in the line.

The results of this analysis are shown below.

Scree plot of Eigen values of sub-factors

The Eigen values suggest that 5 clear factors exist in this data. This is shown in two ways: first by the 'Little Jiffy' criterion (Eigen value > 1), and second, by the marked kink at 5 in the Scree plot.



The pattern matrix of sub-factors

The Pattern Matrix shows how each sub-factor loads on its main factor.

Sub-factor	Factor				
	Affection	Energy	Will	Control	Emotionality
Altruism	1.00				
Trust	0.97				
Support	0.92				
Sociability		0.96			
Vitality		0.94			
Adaptability		0.91			
Independence			0.91		
Determination			0.86		
Confrontation			0.79		
Responsibility				0.99	
Discipline				0.97	
Apprehension					0.90
Tension					0.89

In this table, loadings are sorted in descending order. Loadings below 0.3 have been suppressed for clarity. This information clearly shows that the Facet5 sub-factors load on their expected home factor and contribute little to any other factor. The table below shows the correlations between these factors. The Promax rotation has also managed to keep the Facet5 main factors relatively independent.

Correlations between Facet5 Factors

n = 33333	Will	Energy	Affection	Control	Emotionality
Will	1.0	0.3	-0.3	-0.1	-0.1
Energy		1.0	0.0	-0.1	-0.3
Affection			1.0	0.2	-0.2
Control				1.0	-0.1
Emotionality					1.0

Note that this correlation matrix is almost identical to the previous matrix computed from the original main factor raw scores.

Response pattern and administration method

From the late 1980s to the mid 1990s, Facet5 was used in a traditional, paper and pencil format. Most people took about 20–25 minutes to complete the questionnaire. We then moved to computer based administration and then during 2000 and 2001 Facet5 was web-enabled in response to client-demand. Web administration is considered to offer advantages in terms of increased efficiency, centralised management of the database, ease of updating of libraries and, perhaps most importantly, remote data capture. Most questionnaires are now delivered by the web through any number of platforms.

However, when data is collected by computer as opposed to on paper, questions arise about comparability of results. It is important to determine whether the results are comparable or, if they are not, to know what the effect is.

We have a large amount of data from paper-based responses that has allowed us to identify a 'typical' response pattern to the 106 items in Facet5. Items use a 5-point scale and instructions include:

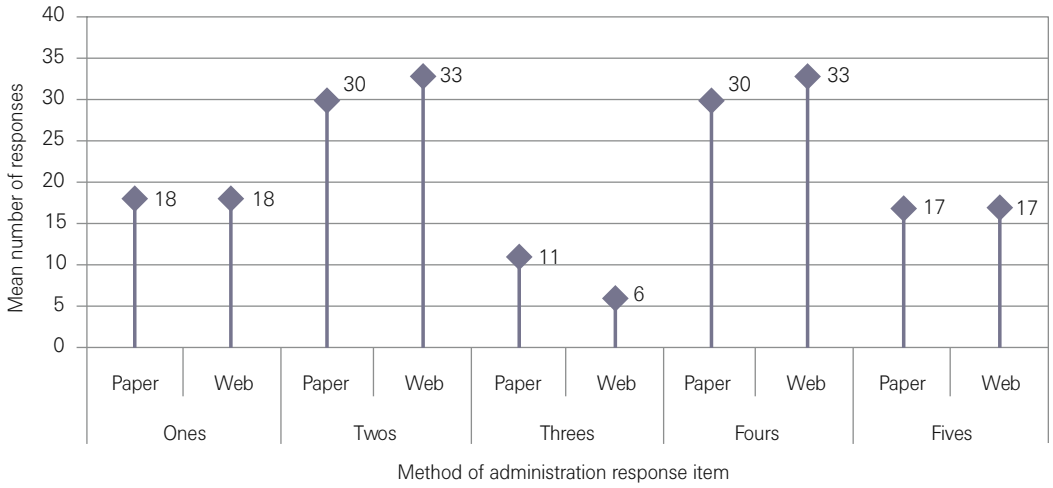
In business most people are prepared to help others	<1> <2> <3> <4> <5>	In business most people are motivated by personal gain

If the statement on the left is most representative of your views, you should circle the figure <1>. If you feel the statement on the right is most typical, you should circle the figure <5>. Try not to fall back on the middle answer (<3>) unless all other answers are completely impossible for you.

If respondents comply with these instructions, we would expect that we would have a bimodal distribution with most answers falling onto 2 and 4. If web-based administration was going to have a significant impact on the response pattern it seems likely that it would show up first here with a pattern which was significantly different. The figure below shows the actual response distribution for both paper and web-based questionnaires. It can be seen that the patterns are broadly identical – none of the differences are statistically significant.

Response patterns and method of administration

Paper N = 12213 Web N = 7584



From this we can be reasonably certain that the shift to web-based administration has not changed the way in which people respond to any significant degree. A second question however was whether a tendency to respond in a particular way was associated with certain personality factors. This was tested next.

Response pattern and Factor scores

From the previous analysis it is clear that most people respond with a bimodal distribution of Facet5 answers. On the five point numbered scale, most people select twos and fours, with fewer ones and fives and even fewer threes. However it would seem possible that some personality styles are more likely to be more extreme in their answers (producing proportionally more ones and fives), while others would be more cautious (produce more threes).

We tested this theory by sorting a large database according to how similar their response patterns were to the 'ideal' pattern of responses. We then grouped them into 20% bands from Band 1 (most similar to ideal) through to Band 5 (most dissimilar to ideal). Finally, we looked at the mean differences for each factor within each band. This table is shown below:

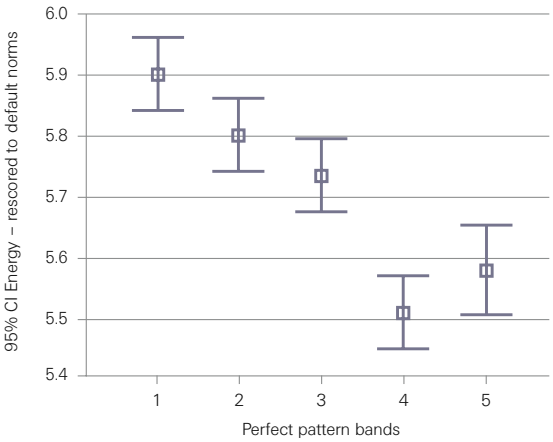
Mean Facet5 scores for Response Pattern bands

Response pattern bands		Will	Energy	Affection	Control	Emotionality
1 ideal	Mean	4.89	5.90	6.74	6.15	5.00
	Std Dev	1.84	1.94	1.88	1.91	1.73
2	Mean	4.89	5.80	6.67	6.10	5.07
	Std Dev	1.88	1.92	1.86	1.88	1.73
3	Mean	4.89	5.74	6.66	6.10	5.19
	Std Dev	1.83	1.91	1.80	1.81	1.73
4	Mean	4.82	5.51	6.59	6.02	5.18
	Std Dev	1.76	1.94	1.83	1.85	1.74
5 less ideal	Mean	4.78	5.58	6.80	6.27	5.07
	Std Dev	1.93	2.08	2.09	1.99	1.90
n = 17797						

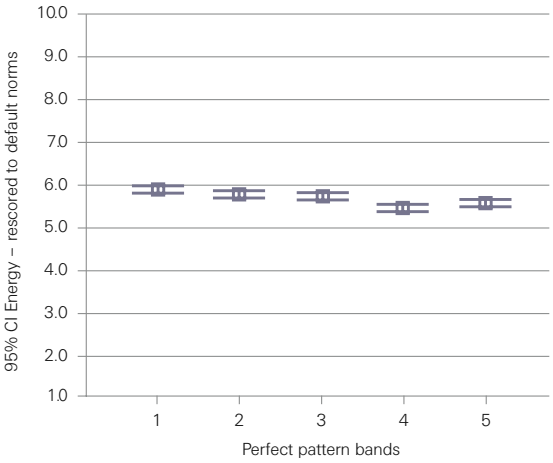
An Analysis of Variance (ANOVA) was computed to see whether any of these mean differences were significant enough to suggest that the degree to which a person responded with extreme answers was related to their personality style. In fact for all five factors there was a 'significant' relationship between response style and personality. However such a large sample base is likely to yield results that appear statistically significant but are in fact of little practical use.
An example can be drawn from the Facet5 Energy factor data.

Energy and response pattern

The mean scores for Energy for each of the 5 bands are shown in the following chart:



This chart shows the mean Energy score for each Pattern Band and the 95% confidence limits (the points that are the limits for 95% of the scores) for each Band. From this chart, visually it appears that there is a relationship between response style and personality. People whose response pattern is 'close to ideal' (Band 1) seem to score slightly higher on Energy than those whose response pattern is 'further from ideal.' In fact the ANOVA is significant at the 0.000 level (df 4 and 18877, F=24.463). But when we look more closely at the figures we can see that it isn't anywhere near as significant as it might at first seem. The chart above has a very compressed vertical scale from Energy scores of 5.4 to 6. If we change the scale to give the complete Sten score range of 1 to 10 we get the chart below.



This looks very different and shows that, regardless of what the statistics say, there's not very much difference across the 5 bands. The proportion of variance in Energy scores that is linked to how 'ideal' the respondent's pattern was very small. Coefficient Eta is designed to evaluate this. Eta squared is the proportion of variance in the dependent variable (Energy) that is explained by differences among groups. The value of Eta squared for Energy across these pattern bands is 0.005 which means that only 0.5% of the variation in Energy scores can be attributed to how the person used the available responses on the scale. And Energy is the most significant of the five factors.

Impression Management (IM)

Impression Management (IM) is an attempt (not necessarily deliberate) to answer questions in a way that will produce a profile that is different from the respondent's 'true' or 'natural' personality. Hogan et al define it as 'from a measurement perspective, (IM) can only be understood as a motivated and significant change from a natural baseline condition of responding'[22].

How important is Impression Management?

As you might expect the issue of Impression Management has been discussed for a long time. It would seem to be a bigger issue when profiles are used for selection than for development but the findings have been quite consistent. Hough and Furnham have summarised the issues well in their contribution to the *Handbook of psychology* published in 2003[23]. Key findings were that:

1. when people are instructed to 'fake' their scores they were able to do so but the differences were larger in a laboratory test than in real life;
2. the level of faking is quite low in selection situations and is not limited to personality tests. Every aspect of the process may be liable to IM including interviews and assessment centres. You would expect people to present themselves in the best possible light. In fact you might be more worried about the person who doesn't try to present well;
3. regardless of whether there has been an attempt to IM, it doesn't seem to affect the predictability of the results.

There are three broad types of IM that might occur:

Denial, defensiveness or suppression

This is an attempt to suppress anything that might be perceived as negative. This can sometimes be identified by specific 'marker' questions or by the proportion of 'No' responses or by some combination of the two. This type of IM is more prominent in selection than development, because a job-candidate is likely to attempt to present the best possible self-portrait.

Indicators of this type of IM are variously termed: Social Desirability, Motivational Distortion or Defensiveness and have their own scales. High scores on these scales are supposed to act as warning bells: the candidate may not be being open or honest.

Suggestibility, faking bad

This second type of IM is the diametric opposite of the first. In this instance a person tends to exaggerate faults or over-admit to possible problems. Such reactions are not uncommon in clinical cases but can also be seen in areas such as Career Counselling or Forensic Psychology.

Such a response pattern is rarely picked up by marker questions. Indeed, a person with this type of response pattern will exhibit very low Social Desirability or Motivational Distortion scores, and are usually interpreted as being 'open' or 'honest'. Examining the proportion of 'Yes' and 'No' responses is more useful, because suggestibility is indicated by a disproportionate number of 'Yes' responses. However, caution is urged against over-simplified interpretations.

Templating

Templating occurs when a respondent has a mental image of 'what is required' or of 'the ideal profile', and tries to adjust their responses accordingly. This is by far the most likely type of IM in selection cases, however typical Social Desirability or Motivational Distortion Scales do little to identify it. Over the years test developers have adopted a number of strategies to minimise the likelihood of IM and to identify it if it does occur. Some techniques include:

- Item ambiguity: this involves wording the items in a way that makes it unclear which factor they load on.
- Neutral or balanced valence[24]: this involves using a scale in which it is difficult to identify which response, representing opposite aspects of a scale, is more desirable. Balanced valence uses pairs of statements that tend to be opposite in meaning and thereby represent opposite ends of a single scale. They may be arranged as separate statements or semantic differentials.
- Marker items: marker items are a group of items that represent 'normal' behaviour. These items will include some that admit negative characteristics on the assumption that most people have some negative elements in their personality. Over-denial of these items flags an attempt by the candidate to appear unrealistically perfect.
- Selective norming: this involves comparing a person's responses only to the responses of people in a similar position. In this way all applicants are expected to have the same degree of motivation to distort and so the effect of IM will be cancelled out. However this assumes that all respondents will apply the same IM strategies to the same degree. It also raises the issue of having to re-norm the profile when the person is hired. How do you then explain that someone's score on a scale was 7 when he or she was an applicant, but it is 9 now that he or she is an employee?

Most test developers will use the first two methods to reduce the effect of IM. Some include the third but the fourth tends to be the domain of people with too much reliance on computer technology. The reason for this is that the collection of multiple norm tables is technically very simple, but it is psychometrically questionable and practically problematic.

When we first developed Facet5, we attempted to address IM in the traditional ways. We created a set of items that produced highly biased response patterns. This suggested that people felt that there was only one way to answer the question. An example of an item that elicited such biased response patterns is: 'I believe in democratic leadership.' Virtually everybody agreed with this notion as a premise and when we created a scale including this and 11 other items, we had a well-balanced scale that had excellent psychometric properties. This putative IM scale correlated positively with Control and Affection, suggesting that people responding in a 'socially desirable' way were attempting to portray themselves as kinder, more responsible and self

disciplined. Another way of interpreting this is that perhaps these respondents actually are that perfect and no distortion exists. It is known for example that Motivational Distortion (MD) and Social Desirability scores tend to correlate with elements of conscientiousness and empathy. In the original 16PF the MD scores correlate with factors A+, C+, F+, G+, H+, L-, M-, O-, Q2+, Q3+ and Q4-. Therefore people with high MD scores are seen as Warm-hearted, Happy-go-lucky, Venturesome, Emotionally Stable, Unperturbed, Relaxed, Conscientious, Practical, Self-sufficient, Controlled, and Trusting. In terms of second order factors broadly aligned with the Big5, these are Extravert, Conscientious and Stable. People who happen to score highly on these scales will tend to have elevated Motivational Distortion scores as well. Other researchers have reported similar results. In his extensive discussion of the impact of Motivational Distortion, Cattell[25] urges great caution in interpreting measures of distortion, describing them as a 'temporary compromise while more basic research proceeds' (page 55). He warns that automatic adjustment using such measures will by definition 'take out real personality variance as well as motivational shift' (page 56).

Response latency analysis as a measure of IM

Clearly, traditional approaches to IM are limited. It may be possible to say that a person is trying hard to look good, but perhaps this is to be expected in a selection setting. Apart from further exploration at interview, there is little guidance as to what to do with the information. Do we infer that the person is lying/faking or are they just 'like that'? If they are faking, are they faking all the scales or just certain scales? It seems likely that people trying to appear more outgoing and forceful will respond to those items they perceive to be related to Outgoingness and Forcefulness. Items relating to Empathy, which they do not see as important, are likely to be subject to much less IM. Traditional approaches to identifying IM will not help much at all.

However, encouragement comes from current deception research. A number of common misconceptions exist about deception, including the notion that deception can be detected through non-verbal indicators such as fidgeting, avoiding eye contact or heightened galvanic skin response. However, some real evidence has come to light. A report in 2001[26] noted that under controlled conditions, a key indicator of an untrue response was the delay in producing the response. Therefore response latency could function to identify deception, and by extension, impression management. Measuring response latency to Facet5 items became possible with web-based administration.

Let us take an example. A candidate for a sales role is keen to get the job and therefore wants to create a good impression. The candidate has read the job advertisement, knows something about the company and has formed a particular opinion about the company from its premises and from the way that she's been received. Then she is faced with a series of questions that under normal circumstances, would appear to have neutral valence. However the situation is not neutral, the candidate will interpret each response against her own interpretation of the company's ideal candidate, and may adjust her response accordingly. When she finds an item that she interprets as 'important' she will think a little more about it and the answer she should give. She may not 'lie' but may easily decide to soften an otherwise strong response. This decision process may be quite quick but will still be slightly slower than her responses to other questions that she does

not perceive as 'loaded'. This pattern of responding would be identical in an online questionnaire format: individuals would take slightly longer to respond to questions they perceived pertinent to their positive impression. This delay can be captured through response latency.

Facet5 is, to our knowledge, the only model that adopts this approach to IM. The concept was first presented to the International Test User's Conference on Internet Based Testing in June 2002[27].

In sum, Facet5 employs item ambiguity and neutral valence to prevent IM, and employs response latency analysis to identify IM if and when it occurs. By decomposing the responses into the Facet5 domains, we can tell not only if IM is present but also which domains are most affected.

Typical response latencies

Facet5 captures the time taken from the moment the item is presented to the time a response is given. However, analysing typical response latencies is not straightforward.
There are a number of issues to consider:

- People differ in their natural response rate. Therefore we must consider the intra-individual response time.
- People do not always complete Facet5 in a single sitting (even though they are encouraged to do so). If the questionnaire is left open the counter will keep running. At present the record is 685853 seconds or nearly eight days.
- People can be interrupted while completing Facet5 by phone calls and messages.
- Facet5 gives the opportunity to skip an item and come back to it later. Which latency do we use?

For these reasons we needed to be selective as to which response latencies we include when trying to understand the way people respond to Facet5. We needed to find a consistent way of handling outliers. To do this we computed the mean response time across all of the 106 items for the UK English data. This mean was distorted by a number of extreme scores.
The results of this can be seen in the table below:

N	29159
Mean	11.56
Median	9.20
Mode	6.34
Standard Deviation	33.51
Minimum	.30
Maximum	4746.17

The effect of extreme scores can be seen in the differences between the mean, median and mode. In this case the mode (the most frequently occurring value) is considered to be the most representative response time.

To reduce the effect of extreme values we tested what would happen if we removed the outliers. The treatment of outliers in data is always contentious but we were guided by Orr, Sackett and Dubois[28] and PsychWiki, 'Dealing with Outliers'[29]. Based on this we declared that all values that were greater than the mode (6.34 seconds) plus 2 standard deviations (2 X 33.51 = 67.02) would be 'outliers' and would be excluded from the analysis. The cut-off value was therefore 73.36 seconds. This produced the following table:

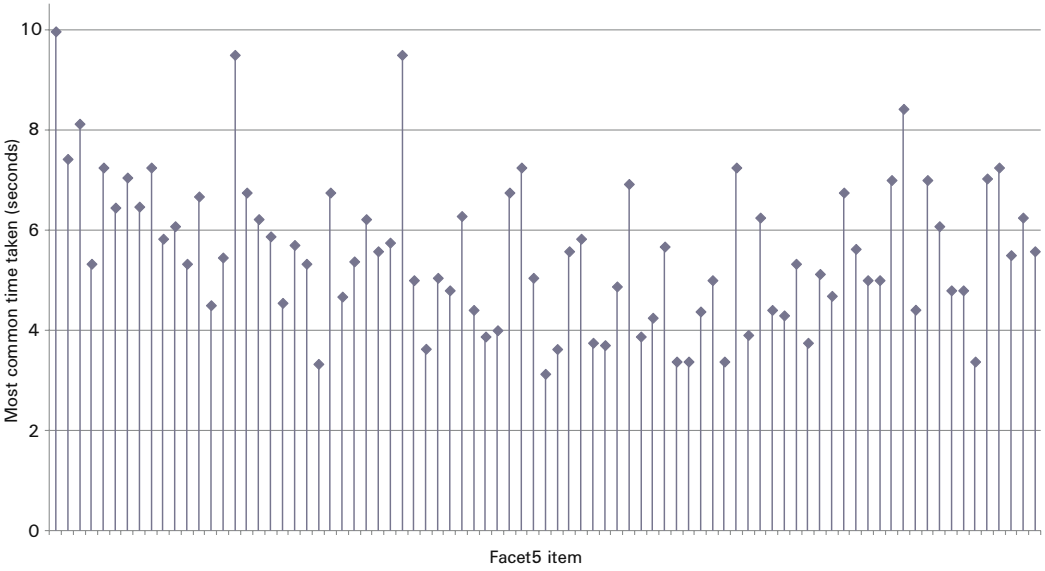
Response times with outliers removed n = 29034

Mean	10.74
Median	9.18
Mode	6.34
Standard Deviation	6.26
Minimum	.28
Maximum	73.34

The most obvious effect of removing the outliers is to bring the mean and median closer together, ultimately bringing both of them closer to the mode, which remains at 6.34. It also reduces the standard deviation dramatically From this we can say that the average response time is between 6 and 9 seconds giving a total questionnaire time of about 10–15 minutes.

The average (modal) response latency for all items is shown below.

Most common response latencies for scored items



As can be seen the modal response latency was a little over 6 seconds. (We chose the mode as opposed to the mean because web based administration is often accompanied by interruptions or distractions which lead to delays on occasional items.) There is sometimes a spike on the first item which we interpret as people getting used to the format, organising coffee etc.
Such spikes also occur on individual records where the respondent is interrupted for some reason.

For any individual there will be a range of response times. Some people will be quick suggesting almost an immediate commitment. Others will be more deliberate and considered.
If there is no attempt at IM the quick and the slow items will be evenly distributed across each of the domains being measured. If on the other hand there is attempted IM then the distribution will be uneven, with longer responses for those domains which are deemed to be 'loaded'.

Response latency and Facet5 scores

A number of Facet5 users have asked whether the tendency to hesitate over some questions is in itself a personality trait. If so then it should be related to the main Facet5 factors.
We looked at this in two stages:

- Are some people just naturally slower to make decisions?
- Do some people naturally have a tendency to try to present themselves in a particular light?

Speed of Response and Facet5 scores #1

The following table shows the relationship between response latencies and the Facet5 main factor raw scores. To get these values we first calculated the response latencies across all 106 items for each person in the above UK English sample (outliers removed as described previously).
We then split the sample into the faster 50% and the slower 50% (median latency = 9.18).
This gave the following results:

Factor	Response time	N	Mean raw score	Standard Deviation	t-value	Significance	Mean difference
Will	Slow	14499	45.69	6.96	9.63	0.000	.81
	Fast	14535	46.50	7.35			
Energy	Slow	14499	51.43	7.78	4.69	0.000	.44
	Fast	14535	51.87	8.23			
Affection	Slow	14499	65.79	8.04	-2.50	0.000	-.24
	Fast	14535	65.55	8.51			
Control	Slow	14499	60.15	9.08	-19.34	0.000	-2.13
	Fast	14535	58.03	9.65			
Emotionality	Slow	14499	47.17	9.39	6.8	0.000	0.78
	Fast	14535	47.95	10.08			

Clearly, there are very small differences between fast and slow decision makers. At first glance this seems to indicate that there is a difference in response time linked to a person's Facet5 score. In particular it appears that people with lower Will, higher Control, slightly higher Affection and slightly lower Energy tend to take longer to make decisions in the Facet5 questionnaire.

Effect size of differences in speed of response

However these t-tests may be misleading. The sample sizes are very large and as a result even small differences can be 'statistically' significant. A better measure is to look at the 'effect size' of these differences: are the differences big enough to be of practical value? We tested this using Cohen's D statistic applied to the means and standard deviations above. The results were:

Factor	Will	Energy	Affection	Control	Emotionality
t – score	0.09	0.17	0.06	0.01	-9.67
Cohen's d	0.11	0.06	0.03	0.23	0.08
Direction of change for slow responders	↓	↓	↑	↑	↓
Effect size	negligible	negligible	negligible	negligible	negligible

On this basis we can see that personality factors have little effect on response latency.

Speed of response and Facet5 scores #2

A second study of English Speaking people (n = 63668) used Sten Scores which had been re-computed to a common norm. The median response time was 9.77 seconds.
The group was divided into 'Fast responders' (median response time is less than 9.77 seconds) and 'Slow responders' (median response time is greater than or equal to 9.77 seconds).
The results were as follows:

Type of response		Will	Energy	Affection	Control	Emotionality
Fast <9.77 secs	Min	1	1	1	1	1
	Max	10	10	10	10	10
	Mean	5.48	5.58	5.67	5.35	5.49
	Std Dev	1.87	1.84	1.87	1.89	1.86
Slow >9.77 secs	Min	1	1	1	1	1
	Max	10	10	10	10	10
	Mean	5.26	5.43	5.72	5.73	5.43
	Std Dev	1.79	1.77	1.78	1.81	1.78
	Cohen's D	.12	.08	0.03	.21	0.03
	Effect size	negligible	negligible	negligible	small	negligible

Here, the differences between fast and slow responders were negligible except for Control where people with lower Control scores tended to respond slightly more quickly. This supports our previous findings.

Do some personalities try hard to make a good impression?

It is also possible that some people naturally try harder to present themselves favourably. If this is true, then we would expect it to be reflected in a link between individual personality factors and a measure of Response Distortion.

If people are deliberately trying to make a good impression then, under the Response Latency hypothesis, there would be an inequality in the time they took over items they thought were important and those which they felt were immaterial. It is important to recognise that this perception will be different for each person and each situation. To test this we computed an index as follows:

- The response latencies for each profile were collected
- The mean and standard deviation of the latencies across all 106 items for each profile were computed
- Outliers were removed (those where the latency is more than 2 SDs above the mean intra-individual latency)
- The remaining latencies were sorted from the fastest to the slowest
- The slowest 15 latencies were selected
- These latencies were sorted according to Facet5 factor

If there was no item which was seen to require more thought than another then we would expect the 15 items to be evenly distributed across the five main factors[30]. The degree to which they did not (ie there was an over representation of one factor) might suggest some form of IM.
To compute this we used the following algorithm:

$$D^2 = \sum_{i=1-5} \left(X_i - 3\right)^2$$

Here, X is the number of the slowest 15 items which load on each factor and i = 1–5 represents the 5 Facet5 factors.

It was thought possible that this index itself might be correlated with personality factors. It may be that some types of people were more prone to hesitate over some questions. We tested this on a small sample of 596 people. These correlations are shown in the table below.

Correlation between Facet5 and amount of distortion in the responses (D^2)

Facet5 Factor	Correlation (significance)
Will	0.064 (.118)
Energy	0.069 (.091)
Affection	-0.060 (.146)
Control	-0.010 (.802)
Emotionality	-0.053 (.200)

n = 596

From this table it is clear that there is no relationship between a respondent's Facet5 scores and how distorted the response latencies are.

Does the situation affect IM?

The situation a person finds themselves in when they are asked to complete Facet5 can vary in its 'valence'. A 'high valence' situation is one where there is a lot of importance attached to the results of the Facet5 profile. A 'low valence' situation is one where the results of the Facet5 profile have little real effect on the person. Recruitment would be high valence, completing Facet5 out of interest would be low valence. We would expect valence to have an impact on impression management and with final Facet5 scores. In order to determine the impact of valence on Facet5 scores, we looked to see whether actual Facet5 scores were different depending on the situation and we also checked to see whether the response latencies differ according to situation.

Does the situation affect Facet5 scores?

From our active database we manually selected a number of people who were known to have completed Facet5 as part of a recruitment exercise. We also selected a similar number who completed it as part of a development process such as team building. We calculated the mean Facet5 scores for the two groups as follows:

Reason for profile		Will	Energy	Affection	Control	Emotionality
Selection (high valence)	N	1693	1693	1693	1693	1693
	Mean	45.47	54.63	67.30	63.27	43.93
	Std Dev	7.01	7.12	8.74	8.80	8.65
Development (low valence)	N	1828	1828	1828	1828	1828
	Mean	47.61	50.89	64.91	56.11	47.95
	Std Dev	7.39	8.21	8.01	9.59	9.86
Significance	Cohen's D	0.3	0.49	0.29	0.78	0.43
	Direction of change	↑	↓	↓	↓	↑
	Effect size	small	medium	small	large	medium

In each case there was a 'significant' difference between the two sets of mean scores. However this was partly an illusion due to the large sample sizes. Cohen's D (Effect size) tells us that the biggest effect was for Control where people who are applying for jobs tend to represent themselves as more Disciplined and Responsible. They also claim they are a little more Outgoing (Energy) and less Anxious and Intense (Emotionality). The impact on Will and Affection was less significant. It should be noted that this effect has been seen in other questionnaires.

Does the situation (valence) affect Response Latencies?

Our second study aimed to determine whether response latency could differentiate between high and low valence situations. We tested this by dividing a sample of people according to the reason that they completed the profile. One group (Selection) completed it as part of a recruitment process and therefore were labelled as 'high valence'. The other group (Development) completed it as part of an investigation into Facet5 itself and were classed as 'low valence'.

The Selection group would be expected to show 'slow item distributions' that were uneven across all five domains as they searched for meaning and importance in the questions. The Development group should have slow item distributions that are more even since they are less concerned with the outcome of the questionnaire. A simple way of assessing this would be to calculate D^2 distributions for the two groups and see if they differ. D^2 was calculated in the same way as before and results were as follows:

	Selection (high valence)	Development (low valence)
N	18	17
Mean DSQ	10.47	7.05
SD of DSQ	3.57	4.19
	T – 2.803 Sig = 0.008 df = 33	
Cohen's D	**0.91 – large effect**	

The t-value tested the significance of the difference between these mean scores and although the sample size was quite small, the differences were highly significant. In fact Cohen's test for Effect Size (Cohen's D) is 0.91 which is defined as a 'large effect'.

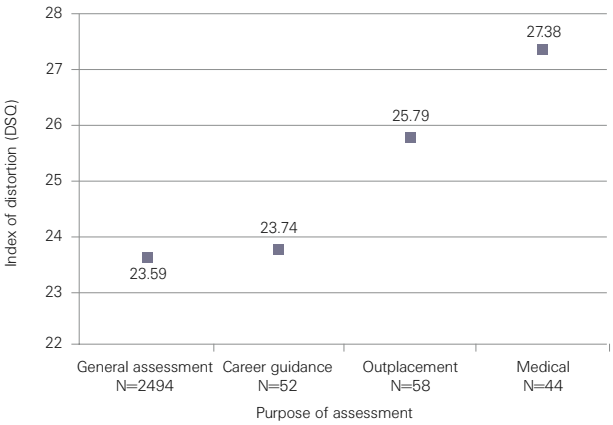
A second study was carried out in 2013 using data collected by a large consultancy. The profiles had been coded according to the reason for the assessment. The codes used were:

Purpose of profile	Description	Expected level of valence
Facet5 feedback N=2494	These people who completed a Facet5 Profile as part of a development process. There was no career or personal decision driven by the results therefore we would assume that Valence would be low.	Low
Career guidance N=52	This group of people had approached the Career Guidance unit of the consultancy for advice on future career direction. Again Valence was assumed to be low.	Low
Outplacement N=58	These people had been made redundant. Profiling was offered as part of the outplacement process. We would assume that valence would be higher for this group since future career direction may be linked to the results of the profile.	A little higher
Medical N=44	This group were required to complete a Facet5 profile as part of court declared evaluation. For example they may have been injured at work or in a motor accident. The purpose of the assessment (of which Facet5 formed part) was to indicate the degree to which their career prospects had been impacted.	Significantly higher

The level of distortion in the response latencies, expressed as a profile match (D2) was calculated in the same way as before and results were as follows:

Purpose of assessment	N	Mean distortion	Std Dev of distortion	SEM of distortion	Minimum distortion	Maximum distortion
General assessment	2494	23.59	8.90	0.18	10.54	52.70
Career guidance	52	23.74	8.30	1.15	10.54	50.00
Outplacement	58	25.79	11.79	1.55	10.54	50.00
Medical	44	27.38	12.71	1.92	10.54	50.00
Total	**2648**	**23.85**	**9.42**	**0.13**	**10.54**	**52.70**

This is shown graphically in the following chart.



In summary this shows that if we take the general assessment score (mean = 23.59, SD =8.90) as the baseline level for distortion then:

1. there is no significant difference between that and career guidance (t=-0.123 df=2544, sig 0.9);
2. there is a slight difference between the baseline and people who were being outplaced (t=-1.846 df=2550, sig 0.06);
3. however, when people were asked to complete Facet5 as part of a critical assessment that might greatly affect their future (medical assessment) they responded in a significantly more distorted way (t=-2.778 df=2536, sig = 0.006). Effect size (Cohen's D) was 0.42 which suggests it is a 'medium effect'.

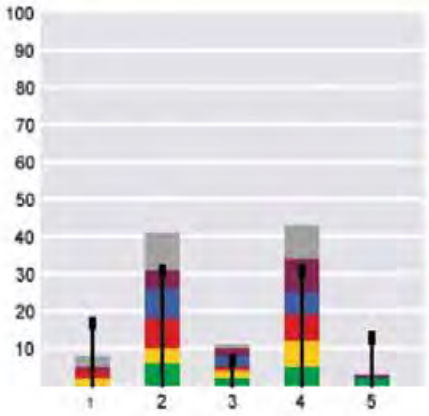
From this data it appears that response latency analysis seems to work. When you have a strong desire to present favourably, your latencies become uneven. We believe that this is because a greater amount of 'cognitive load' is being applied to items perceived to be important and you try to work out what the 'best' answer would be. This would seem to be a promising area for further research.

Analysing, reporting and interpreting response latency

Facet5 reports Impression Management graphically on the 'Statistics' page.
This page presents three analyses:

Response pattern

The response pattern shows the actual distribution of the 106 answers given by a respondent.
It shows this in both a table and a chart as follows:



Factor	Response						Total
	1	2	3	4	5	Reviewer skipped this question	
Will		6	2	5	2		15
Energy	2	4	2	7			15
Affection	2	8	1	7			18
Control		8	3	6			17
Emotionality	1	5	2	9	1		18
Not Used	3	10	1	9			23
Total	8	41	11	43	3	0	106

This chart shows the distribution of responses throughout the Facet5 questionnaire. The vertical bars show the expected pattern based on the responses to thousands of questionnaires completed in the past.

The chart shows the expected distribution (black bars) and the actual distribution (coloured bars). The coloured bars are coded to indicate the Facet5 factor. The table to the right of the chart gives the exact distribution. Note that the items which are not used in scoring are still reported here. That is because when we did not include them many people noticed that the total in the table did not add up to the 106 that they had completed.

A 'perfect' response pattern would be that which matched the 'expected' pattern exactly. It would be bi-modal with most answers being either 2 or 4. It is rare for there to be vastly too many 3s since the Facet5 system is designed to avoid this. Respondents are instructed to try to avoid the 3s unless they really cannot decide. It they answer with more than 18 threes the system will re-present those items and ask the respondent to confirm that they meant to respond with a 3.

Response latency

The response latency for this data is presented as a chart as follows:

Average response time: 4.20 seconds
Fastest: 1.75 seconds
Slowest (excluding outliers): 9.31 seconds

This chart shows the pattern of responses where extra time was taken to answer the questions. Where these responses are focussed on one or two factors it is possible that the overall score on those factors may not be accurate. Other sources of information should be examined to check the accuracy of such scores.

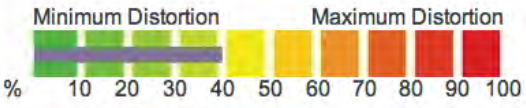
Remember that we produce this chart following these steps:

- Calculate the response times for all items
- Calculate intra-individual mean response time (ie the average response time for that person)
- Calculate the intra-individual standard-deviation of response time
- Remove any responses more than 2 standard deviations above the mean response time
- Sort the remainder from quickest to slowest
- Select the slowest 15
- Count how many of these are associated with each Facet5 factor
- Present as a chart as shown above

Each vertical bar in the chart represents a single item. In this chart you can see that for this respondent, 3 of the 15 items related to Will, 1 to Energy, 3 to Affection, 4 to Control and 4 to Emotionality. The legend shows that the average response time was 4.20 seconds, her quickest was 1.75 seconds and her slowest (with those more than 2 standard deviations above the mean removed) was 9.31 seconds.

Response distortion

The Response Latency chart provides an indication of whether there is an attempt to 'Impression Manage'. However it is difficult to estimate the degree of distortion just by looking at the chart. The Response Distortion bar is designed to help users to see how distorted the latencies are.



This chart shows the level of distortion in the time taken to answer questions relating to a particular Factor. A high level of distortion indicates that Hettie may have been overly cautious with certain types of questions.

There is very little evidence of distortion. 60% of profiles are more distorted. Hettie has been quite open in responding and the profile can be interpreted with confidence.

To produce this bar, Facet5 goes through a number of steps as follows:

Calculate DSQ(resp)

This is the distance between the actual Response Latencies and the 'ideal' Response Latency. The expected value is 3 under each factor so Distance from 3 across all factors is:

$$D^2 = \sum_{i=1-n} \left(X_i - 3\right)^2$$

Calculate DSQ(max)

This is the maximum possible distortion you can get which would be all 15 on one factor (e.g. Will) and zero on all others. When expanded this would look like the formula below:

DSQ(max) = SQRT((15-3)^2)+((0-3) ^2)+((0-3) ^2)+((0-3) ^2)+((0-3)^2) = 13.41641

Compute DSQ(nrm)

This is a normalised value obtained by:

DSQ(nrm) = Divide DSQ(resp) by DSQ(max).

This will give a value between 0 (all Factors have 3 items) through to 1 (1 factor has 15 items and all others have 0).

Compute DSQ(pct)

Multiply DSQ(nrm) by 100 to give a value between 1 and 100. The value of 0 on the chart is labelled 'minimal distortion' and a value of 100 is labelled 'maximum distortion'.

Lookup DSQ in the following table:

Value of DSQ	Label
IF DSQ <= 180.00	This profile is rather distorted. Only 1 in 10 profiles are more distorted so you should check the Response Latency chart for factors where Jennifer has spent a disproportionate amount of time. Interpret the profile with caution. You should check the specific items that caused Jennifer to think longer to see if there is confusion or misunderstanding.
IF DSQ <= 29.95	This profile is a somewhat distorted. Only 20% of profiles are more distorted so you should check the Response Latency chart for factors where Jennifer has spent a disproportionate amount of time.
IF DSQ <= 27.55	This profile is a little distorted. Only 30% of profiles are more distorted so you should check the Response Latency chart for factors where Jennifer has spent a disproportionate amount of time.
IF DSQ <= 25.00	There is only average distortion. 40% of profiles are more distorted so Jennifer is neither more nor less cautious than other people – the profile can be interpreted with confidence.
IF DSQ <= 22.19	There is only average distortion. 50% of profiles are more distorted so Jennifer is neither more nor less cautious than other people – the profile can be interpreted with confidence.
IF DSQ <= 19.67	There is very little evidence of distortion. 60% of profiles are more distorted. Jennifer has been quite open in responding and the profile can be interpreted with confidence.
IF DSQ <= 17.08	There is only minor evidence of distortion. 70% of profiles are more distorted so the profile can be interpreted with confidence.
IF DSQ <= 14.31	There is little evidence of distortion. 80% of profiles are more distorted. Jennifer has been very open in responding and the profile can be interpreted with confidence.
IF DSQ <= 11.33	This profile is not distorted and can be interpreted with confidence

List of slowest items

The Response Pattern and the Response Latency and Distortion charts are printed in the Statistics section of the Facet5 report. The system will also provide a list of the 15 selected 'slow' items sorted by Facet5 factor so you can see what was causing the person to think harder. It looks like this:

15 Slowest Items

These are the items that Hettie took longest over. There can be many reasons for an item taking extra time. Some of the most common include:

Role confusion:
Hettie may have been uncertain whether the question related to the work place or to a different environment.

Item complexity:
Hettie may have been able to see a number of different interpretations in the item and so took time to decide which made best sense.

Role identification:
Hettie was being cautious to not appear too extreme given the expectations of the role.

Questions relating to Will

5	I would prefer to be known as decisive	I would prefer to be known as supportive
21	It is important for a manager to fit in with what the team wants	It is important for a manager to set very clear goals for others
101	If a problem arises with someone at work I act fast and hard	If I have to correct someone I choose the right time and place

Questions relating to Energy

9	I prefer to know what to expect	It is the risk that makes a job interesting

Questions relating to Affection

102	I like to get the benefit of my own efforts	I get great satisfaction out of being able to help other people
1	In business most people are prepared to help others	In business most people are motivated by personal gain
43	I like to do things which benefit other people	When it comes down to it I'll look after myself first

Questions relating to Control

98	It is important to have rules so people know where they stand	Rules stifle individuality
7	There is no need to get too concerned about details	Sloppy workmanship is a sign of an irresponsible attitude
2	I insist on high standards at all times	Different situations require different approaches
3	I'm impatient to get started on new tasks	I can't leave a task till it is finished

Questions relating to Emotionality

72	Once I'm asleep I don't wake up until I have to	I often wake up in the night with something on my mind
40	I remain the same to deal with even under pressure	I am easier to get along with once the pressure's off
4	I prefer to have peace and quiet if I have to concentrate	Noise doesn't interfere with my concentration
20	Having done my best I still can't relax until my review is over	My performance review doesn't worry me, if I've done my best

References and notes

1 Johnson, CE, Wood, R & Blinkhorn, SF, 'Spuriouser and spuriouser: The use of ipsative personality tests', *Journal of Occupational Psychology*, Vol 61 Issue 2, Blackwell Publishing Ltd.

2 The Humm Wadsworth Temperament Scale is the intellectual property of Chandler Macleod Consultants Pty Ltd.

3 Gough, HG & Bradley, P, (1996). CPI Manual. Ed.3. Palo Alto, CA: Consulting Psychologists Press.

4 Hathaway, SR, & McKinley, JC, (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, 10, 249-254.

5 The 16PF has been through many iterations since its original description in 1946. The term '16PF' is a trademark of the Institute for Personality and Ability Testing.

6 Eysenck Personality Questionnaire (EPQ) Eysenck, HJ & Eysenck, SBG, *Manual of the Eysenck Personality Questionnaire*, Hodder and Stoughton, London, 1975

7 Myers, LB, & McCaulley, MH, (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.

8 The DISC model has its origins in the thinking of William H Marston but has been commercialised by numerous organisations and is marketed under many different labels. It is difficult to attribute the current DISC system to any one group but the original thinking appears in Marston, William M. (1928). 'Emotions of Normal People'.

9 Osgood, CE, Suci, G & Tannenbaum, P, (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

10 The idea that Semantic Differentials were faster was developed further in the mid 1970s by Alan Hendrickson and his team working in multi-variate market research. He found that responses to items presented as semantic differentials were completed much faster (response latency of about 4 seconds) than the equivalent item present as a Likert scale. Hendrickson – personal communication.

11 We chose a 5 point scale following advice from Hendrickson which suggested that odd numbered scales, with no middle response, often caused irritation with respondents who were genuinely in the middle.

12 Some people may comment that they do not feel some of the phrases are truly antonymous i.e. not true opposites. This is normal due to the different semantic interpretations people will put on words.

13 Buckley, NL & Williams, RJ. Testing on the web – Response patterns and Impression Management, *Selection & Development Review*, British Psychological Society.

14 Judge, TA, Erez, A, Bono, JE & Thoresen, CJ, 'The core self-evaluations scale: development of a measure', *Personnel Psychology*, Vol 56, Issue 2, Blackwell Publishing Ltd.

15 Costa, PT, & McCrae, RR, (1992b). Professional manual: Revised NEO Personality Inventory (NEO–PI–R) and NEO Five-Factor Inventory (NEO–FFI). Odessa, FL: Psychological Assessment.

16 Comrey, AL, (1970a) The Comrey Personality Scales. San Diego, Educational and Industrial Testing Service

17 Hogan, R & Hogan, J, 1992. *Hogan Personality Inventory manual*. Hogan Assessment Systemsin Tulsa, OK (PO Box 521176, Tulsa 74152) .

18 Hendrickson, AE and White, PO, (1964), Promax: a quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 17: 65–70. doi:10.1111/j.2044-8317.1964.tb00244.x.

19 Professor Eysenck said this in an address to University of NSW psychology students in 1972.

20 Osborne, JW & Costello, AB (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation*, 9(11). Retrieved November 6, 2012 from PAREonline.net/getvn.asp?v=9&n=11.

21 Not all standardisation processes work this way. It is possible to use a cumulative frequency table to work out the scores belonging to the 10th percentile, the 20th percentile and so on. Such methods are particularly useful when the actual distribution is not 'normal'.

22 Hogan, J, Barrett, P and Hogan, R. Personality Measurement, Faking, and Employment Selection, *Journal of Applied Psychology*, 2007, Vol. 92, No. 5, 1270–1285.

23 Hough, LM and Furnham, A, 2003. Use of Personality Variables in Work Settings. *Handbook of psychology*. One: 131–169.

24 The term 'Valence' in psychology means how attractive or unattractive an idea is. Positive valence is something you would like or move towards while negative valence is something you would avoid. The term has probably been borrowed from Chemistry where it refers to the strength of the bonding between two atoms.

25 Cattell, RB, Eber, H, Tatsuoka, MM. 'Handbook for the Sixteen Personality Factor Questionnaire (16 PF)', IPAT, 1988.

26 Bull, R. 'After the Facts', *The Psychologist*, BPS, November 2001.

27 Buckley, NL & Williams RJ, 'Response Patterns and Impression Management', paper presented to the ITC Conference – Winchester UK June 2002-05-8.

28 Orr, JM, Sackett, PR & Dubois, CZ. Outlier detection and treatment in I/O psychology: a survey of researcher beliefs and an empirical illustration, *Personnel Psychology*, 1991, V4, pp473-486.

29 Extracted from www.psychwiki.com/wiki/Dealing_with_Outliers 6 March 2013.

30 Note that this assumption is based on an equal number of items per scale. The fact that Facet5 does not have an exactly equal number of items per scale would change the probability of there being 3 items per scale in the selected 15. This study was conducted under both equal and unequal assumption and there was no noticeable difference in the results.