



**The Orpheus Business Personality Inventory
TEST MANUAL**

Professor John Rust



**UNIVERSITY OF
CAMBRIDGE**
The Psychometrics Centre

Publishers details

Published by the University of Cambridge Psychometrics Centre
Cambridge Judge Business School, Trumpington Street, Cambridge CB2 1AG

Copyright © The Psychometrics Centre

All rights reserved. No part of this publication may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopy, recording of any information storage or retrieval system, without permission in writing from the publisher.

Printed in the United Kingdom

Contents

PART 1: THE OBPI USER MANUAL

Brief description of the OBPI	8
Purpose	8
The OBPI Scales	8
The five personality scales	
The seven integrity scales	
Advantages of the OBPI	9
Instructions for use	10
Before testing begins	
The testing environment	
Establishing and maintaining rapport	
Administering the OBPI	
Keeping records	12
Response audit	12
Standardisation	13
Reliability	14
Interpretation of the OBPI	14

PART 2: THE OBPI TECHNICAL MANUAL

Background of the OBPI	16
Source of items	
Test specification	16
The five OBPI personality scales	
Stability of the big five model of personality	
The seven OBPI integrity scales	
Integrity testing	
The Prudentius model	
The construction of the OBPI scales	26
The personality scales	
The integrity scales	
The pilot study	29
Description of the sample	
Item analysis	
The standardisation study	29
Description of the sample	
Statistical analysis	
Development of response audit variables	
Scale development	
Correlations with sex, age and educational level	
Validity	36
Content validity of the five personality scales of the OBPI	
Content validity of the seven integrity scales of the OBPI	
Construct validity	
Criterion-related validity	
REFERENCES	42

TABLES

Table 1	12
<i>The frequency with which each response audit was obtained in the standardisation sample</i>	
Table 2	13
<i>OBPI scale stanine to percentile equivalents assuming normal distribution</i>	
Table 3	14
<i>Split-half reliability and standard error of measurement obtained from standardisation sample for the five OBPI personality scales and the seven OBPI integrity scales</i>	
Table 4	17
<i>Domain and trait specification for the five OBPI personality scales</i>	
Table 5	18
<i>Names given to the “Big Five” in the literature</i>	
Table 6	21
<i>Some interpretations of specific five factor profiles that have appeared in the literature</i>	
Table 7	26
<i>The classical theory of the Psychomachia of Prudentius together with matching integrity traits from Giotto and the seven OBPI integrity scales</i>	
Table 8	33
<i>Intercorrelations between the OBPI personality scales and correlations with social desirability in the standardisation sample</i>	
Table 9	33
<i>Numbers of items from each of the personality scales used to augment each of the integrity scales</i>	
Table 10	34
<i>The intercorrelations between the integrity scales in the standardisation study</i>	
Table 11	34
<i>Correlations between personality scales and integrity scales</i>	
Table 12	35
<i>Correlations of OBPI variables with age, sex and educational level for the 427 respondents in the standardisation sample</i>	
Table 13	38
<i>The correlations of the seven Prudentius checklist scales with the twelve OBPI scales</i>	
Table 14	38
<i>The partial correlations of the seven Prudentius checklist scales with the twelve OBPI scales, with the social desirability scale (Disclosure) as the partial variate</i>	

Table 15	39
<i>The partial correlations of the seven Prudentius checklist scales with the seven OBPI integrity scales, with the social desirability scale (Disclosure) and the five personality scales as the partial variates</i>	
Table 16	40
<i>Correlations of the five OBPI personality scales with supervisors' ratings</i>	
Table 17	41
<i>Correlations of the seven OBPI integrity scales with supervisors' ratings for 61 security guard personnel</i>	



PART 1: THE OBPI USER MANUAL



Using the OBPI

BRIEF DESCRIPTION OF THE OBPI

The OBPI is a broad-spectrum work-based personality questionnaire. It contains 190 items and generates scores on sixteen scales - five personality scales, seven integrity scales, and four audit scales. The personality scales are Fellowship, Authority, Conformity, Emotion and Detail and are based on the "Big Five" model of personality. The integrity scales are Proficiency, Work-orientation, Patience, Fair-mindedness, Loyalty, Disclosure and Initiative, and are based on the Prudentius model of integrity (Thompson, 1949). The four response audits are Dissimulation, Ambivalence, Despondency and Inattention, and are designed to screen for inappropriate responding.

PURPOSE

The OBPI is intended for wide use within work settings. It can be used for selection, promotion, appraisal or staff development purposes.

THE OBPI SCALES

The five personality scales

The five OBPI personality scales are based on the "Big Five" model of personality, interpreted within the context of work-related behaviours, beliefs, attitudes and interests. The five scales (Fellowship, Authority, Conformity, Emotion and Detail) represent social, organisational, intellectual, emotional and perceptual aspects of personality respectively. Because the personality scales are largely independent of one another, they can additionally be interpreted in combination so that meanings can be assigned to certain patterns of scores. For example, a person with high Fellowship and high Authority (described in the literature as Assertiveness, DeYoung et al, 2016) can be meaningfully differentiated from a person with low Fellowship and low Authority (described in the literature as Submissiveness). Where appropriate, these combinations appear in the interpretative report.

- ▼ **Fellowship** assesses the Big Five trait of extraversion/introversion. High Fellowship scorers are generally happier working with others or in a team. Low Fellowship scorers generally prefer work that requires a degree of independence.
- ▼ **Authority** assesses the Big Five trait of tough- vs. tender-mindedness, sometimes called "agreeableness". High Authority scorers can make tough decisions. Low Authority scorers generally adopt a more co-operative approach.
- ▼ **Conformity** assesses the Big Five trait of "openness-to-experience". High Conformity scorers are likely to prefer traditional ways of doing things and to respect established values. Low Conformity scorers often wish to do things differently, and to seek out alternative solutions to problems.
- ▼ **Emotion** assesses the Big Five trait of neuroticism. High Emotion scorers, while often being of a nervous disposition, are likely to be sensitive to the feelings of others. Low Emotion scorers are likely to be more able to perform under stressful conditions but may lack caution.

- ▼ **Detail** assesses the Big Five trait of conscientiousness. High Detail scorers generally excel at mundane tasks that require particular care, although they may become over-involved in minutiae. Low Detail scorers have less patience for routine tasks and prefer to see the wider view.

The seven integrity scales

The seven integrity scales of the OBPI are designed to assess a person's strengths and weaknesses and are based on a psychological theory of integrity originally attributed to Prudentius. In contrast to the Big Five model, it recognises that most assessments of personality in everyday life are made in terms of the consequences of characteristics or actions for others. Because a weakness in one occupation may be neutral, or even a strength, in another, the integrity scales should only be used where relevant to a work setting. For example, entrepreneurs will often take risks in order to learn from their mistakes, whereas this approach would not be desirable for airline pilots.

- ▼ **Proficiency** assesses the degree of care that is likely to be taken in carrying out a task. It is of relevance to occupations in which mistakes can have particularly severe consequences.
- ▼ **Work-orientation** assesses work ethic. It is of relevance to positions where absenteeism may present a problem, or where staff are required to work long hours or under duress.
- ▼ **Patience** assesses the ability to control aggression in whatever form. It is of relevance to work environments where bullying has been a concern.
- ▼ **Fair-mindedness** assesses fairness in judging the actions of others. It is of relevance to work environments that are beset with strife.
- ▼ **Loyalty** assesses the sense of obedience to company policy. It is of relevance to work situations that necessitate independent action by staff on the organisation's behalf.
- ▼ **Disclosure** is principally composed of social desirability items. Low scores indicate a lack of openness. Bear this in mind when interpreting all other Prudentius scales.
- ▼ **Initiative** assesses a sense of purpose and a forward-looking approach. It is of relevance to organisational settings about to undergo major change.

ADVANTAGES OF THE OBPI

- ▼ Produces a personality profile based on the Big Five model of personality within a work context.
- ▼ Produces seven additional scales based on the Prudentius theory of integrity.
- ▼ Derived from a large established work-based item bank developed in collaboration with several major UK blue chip companies.
- ▼ Constructed to the highest level of psychometric expertise to maximise independence among scales and to minimise response biases resulting from social desirability and acquiescence.
- ▼ Conforms to British Psychological Society and American Psychological Association guidelines on personality test construction and use.
- ▼ Standardised in the UK on a representative sample of employees from a broad range of organisations, occupations, educational levels and ethnic groups

- ▼ Norms produced for individual organisations by arrangement.
- ▼ Produces a reliable and valid psychometric assessment across a broad range of working environments.
- ▼ Quick to administer in an online environment.
- ▼ Incorporates state-of-the-art psychometric procedures in scoring, including within-subject standardisation.
- ▼ Gives full technical and interpretative reports.

INSTRUCTIONS FOR USE

Before testing begins

Test administration should accord with the Company's code of practice. Care must be taken not to infringe equal opportunities and data protection legislation, including the Equalities Act 2010 (Legislation.gov.uk, 2010) and the General Data Protection Regulation and Data Protection Act 2018 (Legislation.gov.uk, 2018).

Candidates should receive a letter in good time before the testing session providing them with clear information about the nature of the assessment, why it is being used, the conditions under which it will take place and the nature of any feedback.

Candidates should also be assured that their test responses will be treated in confidence. They should be told who will have access to the data and for how long it will be kept. Informed consent should be obtained from the candidate before testing begins. This should be in the form of a previously prepared statement which explains the type of test to be administered and its purpose as well as who will have access to the data. It is the responsibility of the test user to ensure that candidates understand the testing procedure.

The assessor should also ensure that all relevant background information from the candidate has been collected and verified and that its storage meets the requirements of privacy legislation. Information on race may be requested in order to enable ethnic monitoring to take place.

The testing environment

The OBPI should be administered in standardised conditions in order to minimise the effects of external factors on test reliability and performance. However, adherence to standard procedures does not mean that the test must be administered in a rigid and unnatural manner. Using a natural conversational tone, encouraging interest in the task and reinforcing the candidate's efforts all contribute to a cohesive and pleasant, though structured, testing situation.

Remember that the physical setting can affect the candidate's performance. To minimise any potential distractions or interference, testing should take place in a quiet, adequately lit, well ventilated room. As a rule, no one other than the candidate should be in the room during testing. If, however, the test is to be administered in a group setting, ensure that all candidates are far enough apart so as not to be able to interfere with each other's performance. Also, when giving any further instructions to individuals take care to do this in a way that does not interfere with others in the room.

Establishing and maintaining rapport

Candidates may arrive for the testing session feeling anxious, suspicious or resentful, all of which may interfere with optimal test performance. It is particularly important, therefore, to establish rapport when the candidates arrive. This may be done by chatting to the candidate and providing reassurances about the nature of the test and the procedures to be employed.

It is essential that the candidate be engaged in the task and be motivated to follow the instructions and complete it as accurately as possible. An administrator, if present, should explain the instructions at the beginning of the test and ensure that the candidate understands their full import. Administrators should be aware of ways in which sabotage can take place. If rapport is properly established and maintained sabotage is much less likely.

It is important for the administrator to maintain a positive attitude to the testing. Not all candidates are likely to be enthusiastic about personality testing, and if the administrator is not able to show enthusiasm any negative attitudes are likely to be reinforced.

If a candidate has a physical impairment which interferes with their ability to complete the test, then the testing environment must be accommodated to the candidate's needs. How this was done, and its likely effect should be included in the report. With visual impairment the items may be read to the candidate.

Various forms of mental impairment can also interfere with test scores. Some forms of mental illness interfere with the ability to concentrate, whereas others may necessitate the administrator reading out the items and sometimes writing down the responses. If this is necessary, these facts and their likely effects should be clearly stated and written in the report.

If a candidate states that he or she does not understand an item, the administrators must use their judgement in deciding how to proceed. If due to the candidate's poor level of education, then testing may have to be abandoned. If, however, it appears that the candidate is merely overcautious, then he or she should be encouraged to endorse their most likely response to the item in question. The development sample for the OBPI included subjects with very little formal education so that problems of this type should be rare.

Although modifications of testing procedures may be necessary, remember that the OBPI was not standardised with such modifications. For example, if sign language or reading of items are necessary, such alterations may have an impact on test scores. Test administrators will have to rely on their professional judgement in evaluating the impact of such modified procedures on test scores.

Administering the OBPI

Candidates should be told what will happen during the testing session, how long it is likely to take, and whether they may leave as soon as they have completed the test. The OBPI is not a timed test, and candidates can take anything between 8 and 30 minutes to complete it. If the test is administered in a group, however, remember that visits to the toilet and early leaving can be disruptive to other candidates.

It is important to emphasise to candidates that they must respond honestly as otherwise their scores will misrepresent them. They may be informed that the OBPI contains a lie detector as well as other filters to detect accuracy of responding.

The instructions should be read to the candidate clearly, and he or she should be encouraged to ask questions before the testing session begins. Candidates should not be encouraged to ask questions once testing has started. Questions about the meaning of items should be discouraged. Any comments made by the administrator may affect the candidate's response and thus bias their scores.

At the end of the testing session, candidates should be thanked for their participation and told about any arrangements for feedback.

KEEPING RECORDS

A register should be kept of all candidates who have attended for testing. The administrator should ensure that the candidate's name or number is clearly and accurately recorded. Completed questionnaire results should be kept in a secure environment and should be accessed only by those with authority to do so. In no circumstances should OBPI results be left lying around an open office. Disposal of any data on paper should be disposed of by shredding and/or incineration.

RESPONSE AUDIT

The OBPI conducts a response audit at the time of scoring. This assigns a value of 0, 1, 2 or 3 to each of four audit measures. Dissimulation is a social desirability score, based on extreme scores on the Disclosure integrity scale, Ambivalence assesses contradiction, and is calculated from a set of similar and opposite items within the questionnaire, Despondency assesses faking bad (relevant to some testing situations, e.g. injury insurance claims), and Inattention, which assesses a tendency to complete the questionnaire in a haphazard fashion. Table 1 shows the frequencies with which each level of audit failure was detected in the standardisation sample.

Table 1: The frequency with which each Response Audit score was obtained in the standardisation sample (N=423)

	% obtaining this score			
	0	1	2	3
Dissimulation	88.9	10.1	7.5	2.9
Ambivalence	89.3	10.7	2.7	0.2
Despondency	95.7	4.3	2.9	1.0
Inattention	94.4	5.6	0.4	0.0

STANDARDISATION

Both personality scale and integrity scale scores are standardised scores, standardised to a stanine scale. Standardisation is a process whereby raw scores are rescaled to make them more meaningful. The standardisation process transforms the raw score of a single respondent based on its relationship to the raw scores of a representative group of respondents, referred to as the norm group. Details of the norm group for OBPI are presented in the Technical Manual. The form of standardisation used by the OBPI is called the “stanine”. The stanine scores on a norm group have a mean of 5 and a standard deviation of 2. If any individual respondent from the norm group has a stanine score of 5, this means that they are at the average for that group on the scale in question. Their score is such that 50% of the other respondents will have obtained a lower score and 50% will have obtained a higher score. If a respondent has a stanine score of 7, then they are scoring one standard deviation above the mean for that scale. This means that 16% of other respondents in the group will have obtained a higher score, and 84% will have obtained a lower score (Equivalent percentages for other scores can be obtained from z tables in any good statistics book or online).

The stanine transformation is carried out using the following equation

$$\text{stanine score} = ((A - B) * 2 / C) + 5$$

where A = the respondent's raw score, B = the population mean of the raw scores on the standardisation sample, and C = the population standard deviation of raw scores on the standardisation sample.

Stanine scores are rounded to the nearest whole number, and values less than 1 or greater than 9 are rounded up or down to 1 or 9 respectively. In interpreting a stanine score it is important to remember the effect of rounding. Thus, a score of 5 represents any score in the range 4.5 to 5.5 before rounding, i.e. it represents the range of scores between -0.25 and +0.25 standard deviations around the mean. Similarly, a stanine score of 6 includes the range from +0.25 to +0.75 standard deviations, and a score of 7 represents the range from +0.75 to +1.25 standard deviations, and so on. Table 2 shows the expected frequency for each stanine score on the assumption of a normal distribution of scores.

Table 2: The OBPI scale stanine to percentile equivalents assuming a normal distribution

Stanine score	s.d. range	percentile equivalents
1	< -1.751	0 - 4
2	-1.75 to -1.251	5 - 10
3	-1.25 to -0.751	11 - 23
4	-0.75 to -0.251	24 - 40
5	-0.25 to +0.249	41 - 59
6	+0.25 to +0.749	60 - 76
7	+0.75 to +1.249	77 - 89
8	+1.25 to +1.749	90 - 95
9	> 1.75	96 - 100

Note: In interpreting percentiles use the upper and lower limits of the range for “and less” and “and more” cumulative percentages respectively. Thus, if a respondent obtains a score of 2, this means that 10% of the population obtain a score at this level or below, and 95% of the population obtain a score at this level or higher. These figures apply to stanine scales which are perfectly normally distributed. Actual frequencies may differ.

RELIABILITY

The split-half reliabilities for the 5 personality and 7 integrity scales from the 423 respondents in the standardisation sample are given in Table 3. For details of the validation of the OBPI you are referred to the Technical Manual in Part 2 of this document.

Table 3: Split-half reliability and standard error of measurement obtained from the standardisation sample (N=423) for the five OBPI personality scales and the seven OBPI integrity scales

	Reliability	Standard error
Fellowship	.73	1.04
Authority	.77	0.96
Conformity	.76	0.98
Emotion	.81	0.87
Detail	.73	1.04
Proficiency	.70	1.10
Work-orientation	.70	1.10
Patience	.73	1.04
Fair-mindedness	.72	1.06
Loyalty	.73	1.04
Disclosure	.76	0.98
Initiative	.72	1.06

Note: The 95% confidence interval for all these scales is 1.96 X the standard error. This rounds to 2 for all stanine scales. Thus, if a respondent obtains a score of 7 or higher on any scale, we can be 95% confident that their score is above the average score of 5. Similarly, with a score of 3 or below we are 95% confident that their score is below the average score of 5 for that scale.

INTERPRETATION OF THE OBPI

The OBPI report summarises the scores on the personality and integrity scales and the four response audits. It should only be interpreted by an appropriately trained professional who is familiar with the details provided in this Manual. The report also provides a personalised interpretative report based on the personality and integrity scales individually and in combination.



PART 2: THE OBPI TECHNICAL MANUAL



BACKGROUND TO THE OBPI

Source of items

The OBPI was derived from a work-based personality item bank containing 420 items drawn from more than 12 customised assessments carried out over a 10-year period. Many of these were broad spectrum assessments which sampled over a wide variety of work-related personality domains. Several different models and approaches were used in these assessments, depending on the specific purpose of each. All candidate items in the item bank had been pre-piloted and many had received earlier modifications as a result, so that the entire bank conformed to acceptable item-analytic criteria.

TEST SPECIFICATION

The OBPI is a purpose-led psychometric test which aims to cover a wide scope of potential applications in respect to test usage in employment settings. Two separate approaches to trait targeting are used. Firstly, the OBPI has been specifically designed to produce a Big Five personality profile. Second, the OBPI has targeted specific assessment problems that are frequently requested by clients. These have been re-specified within a seven-element matrix based on an integrity testing framework, the origin of which has been attributed to Prudentius (Rust, 1996). Two hundred and fifty-three items were selected from the item bank based on their wide applicability. For each target scale, approximately equal numbers of positive and negative items were included in the pilot version.

The five OBPI personality scales

Questionnaires associated with trait models of personality vary enormously in the number of scales they generate. Some target a very few highly stable traits, whereas others generate a plethora of inter-related and relatively unstable measures. This over-abundance of scales has always presented a problem for occupational psychologists in their attempts to make comparisons and choices among different models and instruments. In recent years, one five-factor model has emerged as the favourite and has become the industry standard in terms of making comparisons among instruments with differing numbers of scales. This is the model used in the seminal studies by Barrick & Mount (1991), Tett et al (1991) and Schmidt & Ryan (1993) in summarising the validity of personality tests throughout occupational psychology.

While Thurston (1934), using the factor analytic approach, was probably the first to suggest a specific five factor model, it is more usual to attribute its origin to Fiske (1949). It is he who first noticed that with five factors it was possible to obtain similar factor definitions when different assessment techniques, such as self-ratings, peer-ratings and observer ratings, were used. Tupes and Christal (1961) analysed results from peer ratings, supervisor ratings, teacher ratings and clinical assessments carried out in colleges and in military training and found five strong and recurrent factors across all these domains.

Knowledge of the field was advanced further by Norman (1967) who returned to the origin of personality assessment in psychology - the natural language trait descriptor. This concept owes its origin to the Lexical Hypothesis of Galton (1884) who suggested that any important individual differences between people would have become encoded throughout history in single linguistic terms that would occur in all the world's languages. He identified about 1,000 such personality descriptors. Later, Allport and Odbert (1936) carried out a systematic survey of the English language and listed about 18,000 words in four categories - personal traits, temporary moods or activities, judgements of personal conduct, and capacities and talents. Of these, just under 5,000 "neutral terms designating possible personal traits" received attention by psychologists (among them R B Cattell (1943) who used the list as the basis for his work on the 16PF). Norman argued that earlier studies were flawed as computers, prior to the 1960s, had not been sufficiently powerful to analyse all the descriptors and had therefore had to rely on smaller subsets.

Norman’s initial very extensive work was followed up by Goldberg (1982) who factor analysed 1,710 of Norman’s trait adjectives. In several different studies, he found that, across a variety of samples, there was a very considerable consistency for the five-factor solution, even when different methods of item extraction, rotations and factor numbers were used (Goldberg, 1990). Work by Digman (1990) and by John (1990) also supported the fundamental role of the Big Five model.

Costa and McCrae (1992) argue in favour of the Big Five that the model is supported in four ways (i) the five traits have high stability and are identified using different assessment techniques (e.g. both self-report questionnaires and peer ratings) (ii) they are compatible with a wide variety of psychological theories including psychoanalytic, psychometric and folk-psychological, (iii) they occur in many different cultures and (iv) they have a biological basis. There is good evidence for the first three, and the fourth, while debatable, is not essential to the model.

Within the OBPI, the Big Five model is re-conceptualised as a domain theory of personality, i.e. each of the Big Five factors is unique to its own psychological domain. The five domains are the social, organisational, intellectual, emotional and the perceptual, all of which are essential parts of our psychological life. Thus, individuals live their daily life in a social world comprising their relationships with other people at the personal level. They also live in an organisational world in which their position is determined by hierarchies of social status. They exercise their judgement within an intellectual domain involving the use of reason and knowledge. They are driven moment to moment by their emotions. How they view the world, and what becomes significant in their perceptual field will determine the framework for their actions. Once we know a person’s position with respect to each of these five domains, we have an almost complete description of the functioning of their personality.

The OBPI trait names for each of the five domains are given in Table 4, together with the trait names in most frequent use by other Big Five researchers.

Table 4: Domains and trait specification for the five OBPI personality scales

Domain	OBPI	Big Five trait
Social	Fellowship	Extraversion vs. Introversion
Organisational	Authority	Tough-mindedness vs. Agreeableness,
Intellectual	Conformity	Conventionality vs. Openness
Emotional	Emotion	Neuroticism vs. Confidence
Perceptual	Detail	Impulsivity vs Conscientiousness

Stability of the Big Five model of personality

While considerable evidence has accumulated to show that the five-factor solution tends to be more stable than other numbers of factors, it should always be remembered that this is essentially a consensus. A minority of studies, which may be under different circumstances with different items and different populations, find other solutions provide a better fit, so it is not the case that for every data set the five-factor solution will always emerge. Rather, the five-factor solution is merely the most frequent and ubiquitous. Although it represents the highest level of agreement among experts, there is considerable scope for differing minority views, particularly in populations where special circumstances may apply.

Similarly, the intrinsic nature of each of the Big Five traits also represents a consensus. Again, there is wide agreement concerning the general area covered by each trait. But the specific names given to each trait, and the slant placed on them by the researcher, varies from study to study. Table 5 lists some of the

names that have been suggested in the literature as epitomising each of the Big Five, either by the original author or by subsequent literature reviews by other personality psychologists.

Table 5: Names given to Big Five traits in the literature (Rust, 1996)

Social domain

Extraversion, Surgency, Talkative, Assertive, Energetic, Impulsiveness, Sociability versus Introversion

Organisational domain

Aggression, Tough-mindedness, Moving-against tendency, Inhibition, Narcissism, Hostility, Coronary proneness, Indifference-to-others, Self-centredness, Spitefulness, Jealousy, Authoritarianism, Hostile-non-compliance, Antagonism, Type A behaviour, Thinking versus Abasement, Nurturance, Trust, Self-monitoring, Social interest, Good-natured, Co-operative, Trustful, Altruism, Nurturance, Caring, Emotional-support, Friendly-compliance, Friendliness, Tender-mindedness, Feeling

Intellectual domain

Dogmatism, Sensing versus Openness-to-experience, Creativity, Divergent thinking, Understanding, Change. Sentience, Autonomy, Experience-seeking, Absorption, Flexibility, Achievement-via-independence, Artistic interests, Private Self-consciousness, Culture, Intellectual, Cultured, Polished, Differentiated emotions, Aesthetic sensitivity, Need-for-variety, Unconventional values, Intuition

Emotional domain

Anxiety, Neuroticism versus Emotional stability, Calm, Not neurotic, Not easily upset

Perceptual domain

Achievement, Order, Endurance, Persistence, Competence, Constraint, Prudence, Will-to-achieve, Superego strength, Locus of control, Character, Persistence, Achievement-motive, Dependability, Conscientious, Responsible, Orderly, Perceiving versus Judging

Despite the widespread use of factor analysis in psychometrics there are important differences between the factor-analytic model and more traditional forms of item analysis. The classical personality test is based on the traditional model of assessment (Rust and Golombok, 2009), derived from the idea that a score on a test represents the number of correct responses. It identifies a unique set of items that are used to construct each scale. Factor analysis, on the other hand, constructs traits each of which makes use of all the items in the questionnaire. While it could very reasonably be argued that an improved scale could be made from a more sophisticated use of the factor scores that emerge from factor analysis, rather than insisting that each item be used in only one scale, there are good reasons why the classical model is normally preferred. In test construction psychometricians generally use factor analysis as a tool to advise them in their selection of the best possible items for each scale. Once this selection of items has been made, then it is the scale, and not the factor model, that will need to be defended in terms of standardisation, reliability, validity, bias and other psychometric properties. Factor analytic solutions to data sets may well, for a particular data set, provide factor scores with a better fit than the classical raw sums of item scores. However, with different populations and different circumstances the actual loadings on the factors would be different, and each would need to be defended separately.

One issue which has received attention in the construction of the OBPI is the degree of independence between the Big Five traits. It is often assumed that, because the Big Five are based on a five-factor solution to an orthogonal factor analysis, any scales constructed to measure the Big Five will be independent from each other. This is not the case and independence can only be achieved with considerable difficulty. A failure to recognise the difference between psychometric scales and factor scores underlies much of the confusion concerning the degree of independence of Big Five scores obtained from questionnaires. The five-factor solution on which the Big Five model is based is an orthogonal one - that is, it produces five independent factors. If factor scores were used in place of scales, they would necessarily be independent, if only for the population on which they were based. However, this is achieved by using all the items in the questionnaire five times - each item will have a loading on each of the Big Five although the size of these loadings will depend on the nature of the item in question. Independence in this orthogonal solution is achieved by a very careful balance between the various aspects of all five scales. If only those items which had high loadings on a factor were selected, however, the delicate balancing achieved by counter-balancing the other scales would be lost, and the Big Five scales so derived would no longer be independent of one another. In fact, some of them would have a significant degree of intercorrelation. Attempts to redress this within the factor analysis itself are more likely than not to make matters worse. One solution, for example, would be to use synonyms of certain aspects of each of the Big Five (what Paul Kline has called "bloated specifics"). But this solution would not be the Big Five. This is because the Big Five solution, by its very nature, balances the different aspects of each of its component traits to achieve independence. Simply because the factor analysis generates orthogonal factors it does not follow that the primary traits suggested by each factor will be independent from each other when they are measured separately.

A further complexity of five factor solutions to analyses of personality trait descriptors needs to be considered. This involves the role played by two other elements in all personality questionnaire data sets. These are the response bias effects described as social desirability and acquiescence (Anastasi & Urbina, 1998). Social desirability describes the tendency to fake good or fake bad in questionnaires. Job applicants in occupational settings will often succumb to the temptation to be economical with the truth to some degree. This is not as reprehensible as it may sound as all of us have been encouraged by careers counsellors and others to "make the most of ourselves" in applying for a job. In clinical settings, patients who wish to see a doctor to obtain a sick leave certificate, or to cash in on an insurance policy will often fake bad. The effect of social desirability is ubiquitous, affecting responses to most of the items in one way or another (Paulhus 1984). Consequently, it will always have some effect on the data set and will always influence the factor analysis. This phenomenon is well known to psychometricians, who are constantly struggling to eliminate or neutralise these effects. Unfortunately, it is impossible to do so completely and consequently test users always need to be made aware of how social desirability effects may have influenced results. Acquiescence, the other major form of response bias, represents the tendency of some people to agree with every question or statement, and of others to disagree (Bass, 1955). This effect can also affect every item in the questionnaire and will often emerge as the first factor in a factor analysis. Again, psychometricians have techniques for reducing its influence, usually by balancing the number of positive with the number of negative items for each trait being measured. For the factor analysis itself, however, both phenomena will always be present to some extent.

The effects of response bias on factor structure are not consistent but vary from sample to sample. Thus, students filling in a questionnaire for a research professor are less motivated to lie than are job applicants in a very competitive market (Schmidt & Ryan, 1993). Some forms of bias will be affected by the nature of the job being applied for. Thus, those applying for junior positions would be likely to view different responses as being socially desirable than those looking for managerial responsibility. The perceptual, emotional and organisational domains are particularly likely to be influenced in this way. For example, applicants for sales positions are more likely to bias their responses in the social domain. All forms of response bias affect factor structure, and consequently the nature of the factors will them-

selves vary depending on the nature of the data set and the respondents' motives in agreeing to participate. This is another reason why it makes sense to treat the Big Five structure as a consensus, rather than a scientific discovery concerning the nature of human personality. It describes the most frequent solution under most circumstances. The loadings that occur form a complex family of related items, but one that is bound to be chaotic to some degree and that cannot be completely determined.

Many of the existing questionnaires that measure the Big Five tend to have quite substantial correlations between their scales. These correlations tend to be under-reported, perhaps because the belief that the five-factor solution produces five independent scales (as opposed to five independent factors) is so widespread. While the reason for the intercorrelations between Big Five scales has been explained above, the existence of these intercorrelations is undesirable as one of the main arguments put forward in favour of the five-trait model has been the independence of its domains. This is particularly important if we are to argue, along with Mershon & Gorsuch (1988) that models with smaller numbers of factors are superior, from both a theoretical and a practical point of view, to models with more traits. Cattell (1995) for example has argued that the Big Five model does not address the full complexity of personality, and that a more extensive model, such as his 16PF, is required to dredge this valuable detail.

One argument against having many factors concerns the large degree of association among the traits in these models, which creates a great deal of redundancy within the personality profile. If two traits are highly correlated, they are also conjoined. That is, when one is high then, invariably, so is the other, and vice versa. We could obtain the same information more reliably by simply combining them. Hence, we may just as well have measured only one. With independent traits, on the other hand, our ability to interpret profiles is maximised. A two-factor solution using only Extraversion and Neuroticism, for example, can produce four basic profiles: High Extraversion High Neuroticism, High Extraversion Low Neuroticism, Low Extraversion High Neuroticism and Low Extraversion Low Neuroticism. Eysenck argued that these are related to the four traits identified in classical Greece - the Melancholic, Choleric, Phlegmatic and Sanguine personalities respectively. With three traits we would have more possible combinations, eight in fact if we use all three dimensions. We can also generate further interpretations by taking the three combinations of two traits with the other held constant - producing fourteen interpretations in all (the three traits alone, the three traits in combinations of two and the eight ways in which all three can be combined). With five factors there are, in principle, seventy-four different profiles, each of which could receive an interpretation. Thus, if true independence among the primary scales is to be attained, the five-factor model can provide more information than a questionnaire which claims to assess 70 oblique factors.

Table 6: Interpretations of specific five factor profiles that have appeared in the literature

Dependency:	High Emotion, High Conformity, Low Authority
Social Leadership:	High Fellowship, Low Emotion
Intellectual	High Fellowship, Low Conformity
Submissive	Low Fellowship, Low Authority
Need for recognition	High Emotion, High Fellowship
Defensive attitude	High Emotion, Low Authority
Exhibitionism	High Fellowship, High Authority
Autonomy	Low Emotion, Low Fellowship, Low Conformity
Harm avoidance	High Conformity, Low Authority
Supportiveness	High Fellowship, Low Authority
Achievement	Low Conformity, High Detail
Impulsively	High Authority, Low Conformity
Authoritarian	High Authority, High Conformity

Table 6 shows some common interpretations that have been made for various Big Five profiles. Some of these combinations overlap with trait descriptors that have been put forward for single Big Five traits in Table 5. This should not be particularly surprising as all the Big Five solution provides are generalised clusters that fall within a broadly bounded factor space. The generality, however, is a fuzzy set within a mathematical system that represents data which is essentially chaotic. As soon as an attempt is made to tie down the general characteristics of a domain to specifics at an adjectival level then there must necessarily be some degree of contamination from the remaining four domains. The “true” latent trait of the domain is an abstraction (Loehlin, 1992) which by its very nature eludes precise specification.

One other issue needs to be addressed before we turn to the structure of the OBPI itself - this is the question of whether the latent traits of the Big Five could represent real entities which may be related. Remember that the completely orthogonal nature of the Big Five solution is imposed by the factor analytic technique. However, while it may be the best fitting orthogonal solution, it is possible that the data could be better explained if a less constraining oblique rotation was used - one which allowed the factors to retain some degree of relationship to each other. While this avenue has frequently been explored, it has not been found to be particularly productive. Once the constraint of orthogonality is abandoned, we have lost one of the elements of common ground among our studies - the very one which gave privilege to the five-factor model in the meta-analytic studies in so much current favour. It is not merely a simple question of orthogonal vs oblique, because, short of the complete independence represented by the orthogonal solution there are no grounds for preferring any degree of relationship over any other. To put this another way, if we are no longer specifying targeting a correlation of 0.0 between the factors, it is not always clear whether we ought to target a between-factors correlation of 0.1, or 0.2, or 0.5, nor any other degree of relationship (in practice the degree of relationship between oblique factors is specified by the delta coefficient rather than by correlations). Without the rigor provided by the demand for complete factorial independence there is no fixed standard whereby to judge the appropriateness of any particular factorial solution.

While the exploration of the true nature of the relationship between the five domains represented by the Big Five model is of scientific interest, it is not the issue of most importance to the practitioner and will not be considered further in this Manual. Similarly, we can safely bypass any controversy concerning whether the five traits are biologically based. While it has previously been argued that psychophysiology provides a mediator between personality and biology (Rust, 1975) there is no necessary link between the two (Rust, 1988). Readers who are interested are referred to Costa & McCrae (1992) for a review of supporting evidence for a biological basis for the Big Five. From the practitioner’s point of view all we re-

quire is a convenient structure within which to arrange and summarise our knowledge and experience of personality theory as it applies to occupational settings. This will apply at all levels, from everyday practice to sophisticated meta-analytic evaluations of validity.

The seven OBPI integrity scales

A proper understanding of the OBPI integrity scales demands some familiarity with the literature concerning integrity testing. Integrity tests have been in widespread use in the USA for several decades now. They are controversial and their availability is strictly limited, not so much by statute or by professional code of conduct as by commercial considerations. The larger multinationals that use them tend to develop their integrity tests in-house, while small numbers of instruments that are commercially available are generally restricted to a few major clients of the publisher.

Integrity testing

Most of the major integrity tests available at the time were reviewed in the Tenth Mental Measurement Yearbook (Conoley and Kramer, 1989). At the same time, the American Psychological Association (Goldberg et al, 1991) and the US Congress Office of Technological Assessment (1990) both reviewed the use of integrity tests, and these reports were summarised by Camara and Schneider (1994) in the American Psychologist. In the APA survey, less than half of the publishers of integrity test supplied the information that was requested. Camara and Schneider were concerned that commercial interest had so restricted the dissemination of information concerning the effectiveness of integrity testing that the case for their use was essentially unanswered. They also pointed out that, among the various instruments available, there was little agreement on the behaviours assessed or on a precise definition of integrity. They found the concept of integrity to be overly broad and ill-defined and concluded that there is insufficient evidence to reach clear conclusions regarding the value of integrity testing. Further general criticisms of integrity testing are made by Loevinger (1994), Lykken (see Ones et al, 1996) and Sackett & Wanek (1996).

Sackett et al (1989) classifies integrity tests into two types - (a) overt (also known as "clear purpose tests") and (b) personality based (also known as "disguised purpose tests"). Overt integrity tests include the PSI (London House Press, 1980), the Employee Attitude Inventory (EAI; London House Press, 1982), the Stanton Survey (Klump, 1984), the Reid Report (Reid Psychological Systems, 1951), the Phase II profile (Lousig-Nont, 1987), and the Milby Profile (Miller and Bradley, 1975). Overt integrity tests contain direct questions and biodata items. Personality-based integrity tests, on the other hand, have not been developed solely to predict theft or theft related behaviours. Such tests include the Personal Outlook Inventory (Science Research Associates, 1983), the Personnel Reaction Blank (Gough, 1971), the Employment Inventory of Personnel Decisions (Paajanen, 1985) and the Hogan Personality Inventory (Hogan, 1986). These tests are generally similar in form to any other occupationally based personality test and can be mapped on to the Big Five personality factors in the same manner as other personality tests (Lilienfeld et al, 1995; Ones et al, 1995). A consequence of this is that it is not just integrity testing but the application of personality testing generally to job selection that is under attack by the critics of integrity tests.

The case in favour of integrity testing is made by Deniz Ones and her colleagues (Ones et al, 1993, 1995; Schmidt et al, 1992; Collins & Schmidt, 1993). They report a series of meta-analytic studies which review the evidence for the validity of integrity testing and, on the basis of 650 criteria-related validity coefficients from over 500,000 subjects, conclude not only that the evidence for the validity of integrity tests is substantial but also that the broad construct of integrity is probably as good or better a predictor of overall job performance than any one of the Big Five factors either alone or in combination. In comparison with supervisor's ratings of overall job performance they find an unadjusted validity of .22 for integrity (.41 adjusted for attenuation), which compares favourably with the highest validities reported by Tett et al (1991) in their meta-analysis of occupational personality testing where they found a validity of .22 for the Big Five trait of agreeableness.

O'Bannon et al (1989) list some of the constructs used in integrity tests. These include responsibility, long-term job commitment, consistency, proneness to violence, moral reasoning, hostility, work ethic, dependability, depression and energy level. Ones et al include disciplinary problems, violence on the job, excessive absenteeism and tardiness and theft among target behaviours for integrity tests. Rust (1996), in the handbook for the Giotto integrity test, summarises the target behaviours within a broad-spectrum formulation of integrity based a seven-trait model derived from the psychological theory of the classical scholar Prudentius who published his "Psychomachia" in the fourth century AD.

How does the Prudentius model relate to the Big Five? It will be recalled that proponents of the Big Five model argue that all other personality tests can in principle be reduced to the Big Five factors and, if so, this should also apply to the seven-trait model and, indeed, to integrity testing generally. In fact, Ones et al (1996) argue that many existing integrity tests are to a large extent assessing the Big Five trait of conscientiousness. While there is some force in this position, it overlooks an important difference between integrity tests and most other personality tests. In the construction of most personality tests strenuous efforts are normally made to eliminate the influence of social desirability bias and with some reasonable expectation of success. In integrity tests, on the other hand, the target behaviours are generally such that some degree of social desirability bias is inevitable, and any attempt to eliminate this bias completely would result in a scale which was no longer valid in terms of its original specification. In the assessment of work-orientation, for example, it would be meaningless to require that the scale should be independent of social desirability as work-orientation is by its very nature a desirable characteristic of employees. While the case can be made that personality tests are non-evaluative, the same claim does hold for integrity. Three important points arise from this difference.

Firstly, as pointed out by Tellegen and his colleagues (Almagor et al, 1995) the five-factor model was not, as is so widely believed, based on a factor analysis of all possible natural language personality descriptors. Rather it is based on factor analysis of the natural language personality descriptors that remain after evaluative items have been excluded. According to Tellegen, if these evaluative terms are included in a factor analysis a better fit is obtained by a seven-factor model.

Secondly, while the Big Five model is often believed to be non-evaluative and indeed is promoted as such, the case for this is somewhat weak. Within the Big Five literature, not one positive statement appears to have been made in favour of low conscientiousness. Among employers, a similar lack of enthusiasm seems to be evident for low scorers on agreeableness, and high scorers on neuroticism also find little support. Hence the belief that the Big Five, or indeed any personality test, is non-evaluative is merely a convenience. While the psychological community bends over backwards to find positive things to say about disagreeable neurotics who fail to attend to detail, in practice persons with such scores are treated no more favourably than those who have been administered integrity tests.

Thirdly, response bias effects, such as social desirability and dishonest responding, play an important role not only in integrity testing but in personality testing generally. The lie paradox is well known in philosophy and is best illustrated by Bertrand Russell's example of a piece of paper which has written on both sides "The statement on the other side of this paper is false". Within personality testing the problem is illustrated by the lie item to which respondents are asked to agree or disagree with the statement "I always tell the truth" or perhaps rather more tangentially "It always pays to tell the truth". The impact of this paradox is also demonstrated in items designed to assess social desirability bias, such as "I have never ever made a mistake at work". The consequences for this latter example are worth spelling out. They are (a) a person who is deliberately lying will answer "agree" (b) a person who especially values their honesty will reply "disagree" (c) a person who has been encouraged to present themselves in the best possible light (say, as a result of out-placement counselling) will answer "agree", and (d) a person who does not really care about whether he or she gets the job or not, or who is clinically depressed, will answer "disagree". Essentially, integrity, social desirability and lying are completely confounded in items of this type.

Indeed, as Guastello & Rieke (1991) point out, if we treat social desirability and lying as synonymous then paradoxically many integrity tests have a negative correlation with honesty, and the same is true of the relationship between social desirability and the big-five trait of conscientiousness. Disentangling these effects takes considerable skill on the part of the test constructor.

In constructing mainstream personality tests, every effort should be made to reduce the effects of social desirability. However, this is never completely possible. The Big Five trait of conscientiousness always retains some degree of social desirability bias, as does the trait of neuroticism. In fact, for all traits social desirability may have an effect under specific circumstances. Extraversion scores, for example, are likely to be affected in different ways if the job applied for involves either working in a team or independently. Generally, the more relevant the trait to the job, the more likely the scores are to be biased. With integrity tests a different approach is to be recommended. While the reduction of obvious or unnecessary sources of social desirability bias is still important, it is recognised that it would be not only impossible but also ill-advised to attempt to eliminate the effects of social desirability completely. Given the nature of the traits being assessed, some degree of association with social desirability is inevitable. Instead, social desirability should be assessed in addition to the trait of interest by a specific scale designed for this purpose. Interpretation can then be based on a judgement of both scores taken in combination.

How can the problem of social desirability be tackled? In many ways the problem is the same as that faced by an interviewer when dealing with an accomplished liar, and there are two courses of action that can be recommended. Firstly, the interviewer can ask indirect questions. This has the advantage that the correct answer is not so obvious and therefore lying less likely to affect the response, but the disadvantage is that the information gained will be somewhat tangential to the main points of interest. Secondly the interviewer could ask direct questions and try to assess the extent of confidence that can realistically be placed in the answers. Where this works it will be very successful, however where this does not work the interviewer may not only be left in the dark but may also make the wrong appointment. The first of these approaches is the one used by the OBPI personality scales. The second approach is used by the seven OBPI integrity scales. The skilled interviewer is not restricted in the type of questions (either direct or indirect) that can be asked and will frequently utilise both in combination.

In a recent joint paper by Deniz Ones and David Lykken, an exchange of letters concerning integrity testing is reported which provides a useful summary of the debate so far (Ones et al, 1996). Lykken is well known for his criticism of the use of psychophysiological lie detectors in US industry and played a part in their eventual outlawing. He now wishes to extend many of the same criticisms to the integrity tests which he sees as having replaced the lie detector and as having many of the same characteristics. Lykken argues that (1) the database used by Ones and her colleagues for meta-analysis includes studies employing polygraph screening tests or admissions of previous dishonesty as criteria, (2) that if integrity tests are used for selection then large numbers of people may be permanently denied jobs, (3) that few procedures are less scientific or publicly accountable, (4) that validities stated are from publishers handbooks and have not been subject to peer review by the academic community, and (5) that integrity testing is likely to do considerable harm and should be monitored and regulated.

Ones and her colleagues reply that (1) Studies including polygraph admissions had been specifically excluded from their meta-analysis. They had also examined separately the studies which included admissions of previous dishonesty and those which did not, and found that the results were the same in both cases (2) that all selection procedures had the potential for excluding a large fraction of job applicants as this was the whole point of selection, (3) that there is scientific evidence for integrity testing which is publicly available in the scientific literature, (4) that they had analysed data originating from publishers handbooks and from academic studies separately and again found no difference, and (5) that any argument against integrity testing on the grounds of its impact would apply equally to personality and ability testing in occupational settings.

In conclusion, we can see that integrity testing while remaining controversial, raises questions which are not different in kind from those which are already familiar from the literature on ability and personality testing. The anxieties that are generated in all these areas arise from the important role that job selection (and by default also job rejection) plays in the lives of all of us. But this is a by-product of the selection process in general and is not a fault of the tests in themselves. It should be a requirement for any selection process that it be reliable, valid and unbiased, and these psychometric criteria should be applied to the evaluation of all selection techniques, whether they be by interview, by questionnaire, by test of knowledge, by qualification, or indeed by any technique whatsoever. The value of any technique will stand or fall based on publicly accountable scientific evidence. There can be no doubt that integrity tests address directly some of the major questions of interest to an employer in the selection process. There is currently an increasing interest in the utilisation of personality tests, particularly those based on the Big Five, to assess integrity. However, the use of more focused scales to assess particular attitudes and behaviours necessarily remains of considerable interest to all involved in personnel assessment. Indeed, it is frequently the identification of those attitudes and behaviours addressed by integrity tests, such long-term commitment, laziness and absenteeism, which is the primary argument in favour of using personality tests at all. In the OBPI integrity scales an attempt is made to address some of these issues directly.

The Prudentius model

It is the received wisdom that the theories which underpin the scales of personality questionnaires are of two types - theoretical and statistical. Thus Eysenck's theory is seen as theoretical as it draws its inspiration from the four humors of Ancient Greek science, while Cattell's theory is seen as statistical as it was derived from a factor analysis of natural language personality descriptors. However, the distinction is not so clear cut. Factor analysis is widely used in the construction of scales to assess theoretical traits, and the folk psychology implicit in the natural language descriptors themselves has a major impact on the outcome of factor analyses of "theoretical" data sets.

Before Galton and prior to the advent of mathematical statistics, natural language personality descriptors had already been a subject of interest to classical scholars and are considered in many learned texts from the classical period. Probably the most influential of these was the *Psychomachia* of Prudentius in the fourth century AD. This model was later adapted by theologians and thinkers throughout Christianity and Islam, and the fundamental psychological concepts involved have influenced much of our thinking today.

Prudentius established the convention of opposing virtues and vices representing bipolar personality characteristics which he believed were fundamental to human nature. These were Faith and Idolatry, Chastity and Lust, Patience and Anger, Humility and Pride, Temperance and Gluttony, and Charity and Avarice. Dante, in the *Purgatory* (Sayers, 1955), used this framework as the model for his "Deadly Sins" - Pride, Envy, Anger, Sloth, Avarice, Greed and Lust. Giotto, in his depiction of the vices and virtues at the Arena Chapel in Padua, Italy in the 13th Century AD described the characteristics as Justice/Injustice, Hope/Despair, Charity/Envy, Faith/Idolatry, Temperance/Anger, Fortitude/Inconstancy and Prudence/Folly. Using any computerised thesaurus to identify the synonyms and antonyms of these concepts in modern English it is easy to identify the central role played by them in the personality descriptors used in everyday language. See the *Giotto Handbook* (Rust, 1996) for a more comprehensive coverage of the relationship between the Prudentius model and integrity testing. The relationship between the Prudentius model, Giotto and the OBPI integrity scales is shown in Table 7

Table 7: The classical theory of the Psychomachia of Prudentius together with the matching integrity traits from Giotto and the seven OBPI integrity scales

Classical virtues/vices	Giotto Integrity trait	OBPI integrity scale
Prudence/Folly	Carelessness	Competence
Fortitude/Inconstancy	Absenteeism	Work-orientation
Temperance/Anger	Hostility	Patience
Justice/Injustice	Subversion	Fair-mindedness
Faith/Idolatry	Disloyalty	Loyalty
Charity/Envy	Greed	Disclosure*
Hope/Despair	Inertia	Initiative

Note: The OBPI integrity trait of Disclosure maps on to social desirability rather than to the integrity trait of Greed. The OBPI integrity scales assess the reverse of their corresponding integrity trait.

THE CONSTRUCTION OF THE OBPI SCALES

The personality scales

The OBPI scales were constructed progressively and incrementally from core attributes of their traits which had been predefined for items with the use of markers (Rust & Golombok, 2009; Goldberg, 1992; Peabody, 1987; Peabody & Goldberg, 1989). As the scales emerged, they were carefully balanced for positive and negative items, to avoid acquiescence effects (Knowles, 1963; Paulhus, 1991). Positive items are those which are summed directly into the scale, negative items are reversed before being summed into the scale (e.g. “I am sociable” would be a positive item, and “I am not sociable” would be a negative item for a scale of sociability). Scale intercorrelations were also continuously monitored as well as item/total correlations and alpha coefficients to ensure both breadth and diversity in trait coverage. The five personality scales were modelled around the Big Five template shown in Table 4 with the additional criterion that no two Big Five traits should have intercorrelations above 0.3, and that the correlation with a previously constructed interim social desirability scale should also be as low as possible. The seven integrity scales were modelled in the same way around the Prudentius template (see Table 7). While neither of the two previous restrictions were applied directly to the construction of the integrity scales, the requirement that each item should be associated with only one integrity scale in terms of both reliability and validity kept artificial correlations between the seven integrity scales to a minimum.

Given this development process, how might we expect the five personality scales from the OBPI to differ from other questionnaires that have utilised the Big Five model? Firstly, the requirement that scales should be as independent as possible from lie and social desirability effects will help to identify any meritorious personality characteristics which may be associated with individuals who obtain “undesirable” scores on a scale. While it is widely believed that the Big Five model is non-evaluative, inspection of the scales shown in Table 2 will demonstrate that contamination with social desirability is rampant throughout the Big Five literature. Given the labels used in many questionnaires, there seems to be little support for the claim that the Big Five structure is non-judgmental. If, however, social desirability bias can be reduced further, then it is to be hoped that, in future, Table 2 will be replaced by a more balanced picture.

The requirement that correlations between each of the Big Five scales be less than 0.3 addresses a set of common problems. One of these has been the degree of concordance between extraversion and openness-to-experience. While extraversion almost invariably includes a social aspect, different extraversion scales vary in the extent to which impulsiveness items are included. The disagreement owes its origin to Eysenck (1947), who, in the earliest versions of the EPI (the Maudsley Personality Questionnaire), focused

on the independence of extraversion and neuroticism. Two factor solutions always provide the simplest model for the use of factor analysis in personality test construction. While the nature of the two factors that will emerge from this model will always depend to some extent on which items are in the pool, it has long been known that for any wide scale sampling of natural language trait descriptors the most likely two factors resemble extraversion and neuroticism. Either is likely to emerge as the first factor, and once it has been identified it is relatively easy to construct an independent second scale from prudent balancing across the remaining items in the pool.

Unfortunately, as soon as the number of factors is increased beyond two, this structure is liable to disintegrate. Eysenck himself was only able to generate an independent third factor of psychoticism by removing impulsivity items from his earlier extraversion scale (Claridge & Birchall 1978, Rust, 1975,87,88,89). When the number of factors is further increased from three to five, impulsiveness tends to transfer to openness-to-experience rather than to psychoticism. Within the OBPI, the constraint on the size of the correlation between Conformity (Openness-to-experience) and Fellowship (Extraversion) has produced both Fellowship and Conformity scales less contaminated by impulsivity, which is now picked up by the high Fellowship, low Conformity profile.

Another common confound has been between the Big Five traits of Detail (conscientiousness) and Authority (agreeableness). This has largely come about because almost all the development work on the Big Five was carried out in academic institutions using students as respondents. The friendly and conscientious student is highly valued by staff and fellow students alike and the benefit of this trait is seen as self evident. However, although people low on “agreeableness” may not make ideal students, there are a very large number of them - and many are very successful in their careers. The experience of many occupational psychologists is that those with low scores on agreeableness and/or conscientiousness are not always the ne’er-do-wells that have been implied. Successful senior managers, and human resource managers, very frequently have this profile, so that its positive qualities do need to be addressed. By focusing on low correlations between Authority and Detail, and by reducing the mediating effect of social desirability, the OBPI has achieved a good degree of independence between these traits. Many of the items for these scales which correlated more highly with social desirability were transferred to the item pool from which the seven integrity scales were constructed. This has led to an emphasis on the previously obscured merits of low scores on agreeableness and on conscientiousness in the OBPI Authority and Detail scales respectively.

Within the OBPI structure, an “agreeable” and altruistic outlook at work is seen as more appropriate for high performing junior staff and not necessarily a suitable quality for those who must direct the activities of others. People who have heavy managerial responsibilities may often see themselves as tough-minded or even “disagreeable”. With the conceptualisation of “disagreeable” respondents as high authority it is possible to give recognition to their positive qualities, their ability to make tough decisions and their will to succeed. The distinction between the Social and the Organisational domains also helps to disentangle the social quality of agreeableness which otherwise tends to relate its friendliness aspect to extraversion. By treating “agreeableness” as a characteristic of those possessing less onerous responsibilities it is possible to distinguish well-meaning and kind but introverted individuals from their more demonstrative and ostensibly helpful colleagues.

There has also been a tendency within working populations for low agreeableness (tough-mindedness) to correlate with high openness-to-experience. This is very likely an artefact introduced by the relationship between status and education. High Authority individuals are generally more senior and more educated - often graduates. Such people tend to be more intelligent, which in turn relates to lower scores on Conformity. The relationship between openness-to-experience and intellect in the form of divergent thinking has been widely explored in the literature. However, it is a mistake to treat openness-to-experience as merely a surrogate for intelligence. In fact, the relationship with IQ scores, although usually present, is

not particularly straightforward. By targeting a lower correlation with Authority and by placing the trait in the intellectual domain, the OBPI deliberately emphasises its relationship to the process of reasoning rather than its behavioural attributes. The concept assessed by Conformity should be seen in the context of epistemology and the philosophical theory of knowledge rather than the more prosaic psychological concepts of intelligence. Low Conformity scorers aim to change their own ways of thinking; high Conformity scorers are more willing to take things on faith.

The OBPI interprets conscientiousness as attention to detail, and thus places its trait of Detail within the perceptual domain concerned with how and where we focus our attention. While the relationship between conscientiousness and attention to detail is well documented, relatively little attention has been paid to the attributes of low scorers on this trait as can be seen from the relative paucity of descriptors it receives in Table 2. Where trait descriptors for low Detail scores have been negative, they have invariably been portrayed as extremely undesirable, using such terms as unreliable, lazy, hedonistic, and lacking in self control and moral principles. Despite this, persons with scores at this end of the Detail scale often seem to be fully functioning and successful individuals! It is hoped that the OBPI framework will address this injustice and emphasise their ability to “pan out” from the narrow focus of the highly Conscientious individuals. Low Detail scorers may be more able to use their intuitions and more able to make use of a gestalt style of perception, recognising that the whole is often more than the sum of its parts. They epitomise the impressionist in contrast to the photographic mode of perception.

The integrity scales

The Giotto integrity test was specifically designed to address statistical problems involved in the assessment of integrity, the non-linearity of many of the fundamental concepts and of social desirability itself. The seven OBPI integrity scales, however, are constructed to the classical psychometric model which assumes linearity. An analysis of cross-correlation between Giotto and the OBPI items showed that the focal point of the resultant discrepancy was the Greed dimension which mapped directly onto the interim OBPI social desirability scale, itself highly related to social desirability. This interim social desirability scale has therefore been firmed up as one of the seven integrity scales described as Disclosure. It represents, metaphorically, the eye of the storm, or the black hole at the centre of the non-linearity introduced by the very existence of lying and social desirability bias at the heart of personality testing. Any information which might be contained in a scale to assess honesty is, by its very nature, immediately lost in the lies that are told. It is usual in personality test development to avoid this vortex as far as possible by excluding items which correlate too highly with social desirability. This is rarely done with complete success, and most personality questionnaires have a few scales that stray dangerously close or fail to report social desirability scale correlations for those scales for which this influence is suspect. The reason is obvious, the more pertinent a particular trait to a particular job the more likely respondents will feel tempted to endorse items perceived as representing what is required. Thus, paradoxically it is those very traits that are vulnerable to the distorting effect of lying that are of most interest. Unlike the five personality scales, the seven OBPI integrity scales explore the space closer to this vortex in mapping onto the Giotto framework.

The OBPI integrity scales were targeted onto the Rust (1996) integrity framework (see Table 7). Sets of opposing pairs of items from each end of the bipolar trait in question were selected from the pre-piloted items in the item bank with known psychometric properties. Factor analysis of appropriate items for each scale was carried out, and the four items (two positively loaded and two negatively loaded) used to form an interim scale which was then correlated with the remaining pilot items to generate a pool of potential items for scale accumulation. Items for the Disclosure scale were selected from the pool of social desirability items within the item bank which also formed an interim social desirability scale for the construction of the five personality scales. Only items of relevance to work-related settings were sampled. As all items within the bank had been pre-piloted in terms of difficulty level, no item exclusion on the

grounds of extremity of difficulty value was anticipated. Scale construction for all traits proceeded in parallel using the discrimination index from within the classical item analytic procedures on the appropriate face and construct valid items, retaining a balanced number of positive and negative items as scale length grew. The growth of the scales was not strongly constrained in that intercorrelations between scales or correlations with the interim social desirability scale were not among the item selection criteria. This said, the face validity of items was never ignored. Criteria that were used related to divergence among items (items should not be too like each other and certainly never synonymous) and homogeneity of item total correlations.

THE PILOT STUDY

Description of the sample

The 253-item pilot version of the OBPI was administered to 274 employees in a variety of occupations, ranging from junior technical and clerical staff to senior managers and professionals. A broad sample of work settings were sampled, including a major automobile manufacturer, a major police force, an industrial petro-chemical concern, a retail chain and a city financial institution.

Item analysis

Item analysis of these data, together with those already available in the item bank, gave 12 scales. The technique for scale construction followed a procedure developed by Rust (1976,1988,1989, 2009). Each scale had balanced number of positive and negative items. The five personality scales of Fellowship, Authority, Conformity, Emotion and Detail all had intercorrelations of less than 0.3, and each had a correlation less than 0.3 with an interim social desirability scale specifically constructed for this purpose from social desirability items that had been included in the pilot. Once the 5 personality scales had been identified, seven integrity scales were assembled from the remaining pilot items. One of these (Disclosure) was assembled from the interim social desirability scale. The final version of the OBPI consisted of 190 items.

THE STANDARDISATION STUDY

Description of the sample

The OBPI was administered to 423 respondents in a variety of occupations within over 20 companies in the UK. There were, for example, 20 teachers and trainers, 13 accountants, 46 security staff, 37 managers 18 drivers, 25 engineers and scientists, 23 sales and marketing staff, 6 police officers, 10 insurance claims negotiators, 42 secretaries and clerks, 21 human resource personnel, 17 insurance underwriters, and so on. The sample included 275 men and 138 women, and the mean age of the sample was 30.67 years with a standard deviation of 11.01. Ages ranged between 16 and 62 years, including 40 16/17-year olds and 17 aged 50+. The educational level of the respondents ranged from 22 with no qualifications, through 76 with GCSE, City and Guilds etc., 69 with GCE O level, 80 with GCE A level or ONC, 90 with University degrees or equivalent, 25 with Masters degrees or higher professional qualifications to 5 PhDs. There were 30 respondents who did not classify themselves as “white” on the ethnic monitoring question. This included 10 Asian, 14 Black and 6 Other, and represented the approximate proportion of these minority groups within the UK working population. The population sampled for the standardisation study was broadly sampled and, with certain caveats, can be taken as representative of the UK working population. The accumulated data from all the respondents in the standardisation study provide the general population norms for the OBPI. These are used to produce the standardised stanine scores which are used in the interpretative report.

Statistical Analysis

The data were standardised within each respondent (Rust & Golombok, 2009). That is, for each completed questionnaire the mean and the standard deviation of all item responses was calculated, and the standard score for each item was computed using the z score formula $z=(x-\text{mean})/\text{s.d.}$ This was carried out in order to eliminate response bias effects resulting from acquiescence. Item analysis, reliability analysis and validation were conducted with data that had been pre-transformed in this way.

Development of response audit variables

Four response audit (screening) scales were derived from the standardisation data. These were:

- ▼ Dissimulation
- ▼ Ambivalence
- ▼ Despondency
- ▼ Inattention

Scores on all response audit scales range from 0 through 3, with most respondents scoring 0 on all audits.

Dissimulation

The Dissimulation screen was derived from the Disclosure scale (after standardisation but before rounding) as follows:

Percentage obtaining this Dissimulation screen in the standardisation sample

Disclosure < 1.500 s.d.	Dissimulation = 0	88.9
Disclosure > 1.500 s.d. and < 1.875	Dissimulation = 1	3.6
Disclosure > 1.875 s.d. and < 2.250	Dissimulation = 2	4.6
Disclosure > 2.250 s.d.	Dissimulation = 3	2.9

A Dissimulation score of 3 indicates that the respondent is almost certainly lying deliberately, a Dissimulation score of 2 indicates a very high probability that the respondent is lying, and a Dissimulation score of 1 suggests that the possibility that the respondent is lying should always be considered in interpreting his/her scores.

The scale on which the Dissimulation screen is based (Disclosure) was itself derived from the pilot social desirability scale which is discussed in more detail later. The psychological literature on social desirability scales shows that lying has a very complex (and generally non-linear) relationship with personality test scores. Respondents in most real-life situations are expected to present themselves in a good light, and indeed are advised to do so by career consultants. Some inflating of scores, particularly on socially desirable traits, is therefore to be expected. The point at which this is treated as dishonesty, rather than a wish to present the best of oneself, is likely to vary from situation to situation, and test users should consider other factors in the respondent's background when interpreting their scores. These will include the circumstances of testing and the background of the client.

Ambivalence

The Ambivalence screen was calculated from raw (not standardised within subject) data from a count of the number of contradictions from the 12 most strongly opposed set of items. A contradiction was identified after a pooling of strongly disagree with disagree and of strongly agree with agree. It was judged

to have occurred if the meaning of the two responses were to some extent in contradiction. For six of the items, contradiction consisted of agreeing in one instance and disagreeing in another instance to two similar items. For the remaining six items, contradiction consisted of either agreeing to both or disagreeing with both of two contradictory items. All item pairs in the first case had correlations of greater than 0.44 in the standardisation sample. All item pairs in the second case had correlations of less than -0.35.

The mean raw score (out of 12) was 3.34 with a standard deviation of 1.62. Out of 401 respondents, 10 had a raw score of zero, 37 had a raw score of one, 88 had a raw score of two, 93 had a raw score of three, 74 had a raw score of four, 56 had a raw score of five, 32 had a raw score of six, 10 had a raw score of seven, and 1 had a raw score of eight. The reason that these raw frequencies appear quite high is that items that duplicated unduly or that were direct contradictions were eliminated at the test development stage. The item pair 90-142 was the exception, and these had an intercorrelation of -0.70. The other pairs in the Ambivalence screen, however, while having much in common, were not direct semantic contradictions but rather were ones that were endorsed rarely in a particular combination by respondents in the standardisation study.

The Ambivalence screen was derived as follows from these raw numbers of contradictory statements.:

Raw contradictions	Percentage obtaining this Ambivalence score in the standardisation sample	
5 or less	Ambivalence = 0	89.3
6	Ambivalence = 1	8.0
7	Ambivalence = 2	2.5
8 or more	Ambivalence = 3	0.2

Any person who obtains a high Ambivalence score is unlikely to have been paying much attention to the content of the items and may well have been approaching randomness in their responding. This is more probably the case the higher the Ambivalence score.

It should be noted that the expected score for a “genuine” random responder on Ambivalence is 1 (6 out of 12 on the binomial distribution), and there is therefore an about 40% chance that such a random pattern will fail to be detected by the Ambivalence screen. This situation bears comparison with that found when contrasting type 1 and type 2 errors in statistical hypothesis testing. That is, the screen will generate fewer false positives to the hypothesis “Is this pattern of responding random?” when it is false, than false negatives when the hypothesis is true.

Despondency

The Despondency audit screen was derived from the Disclosure scale (after standardisation but before rounding) as follows:

Percentage obtaining this Despondency score in the standardisation sample		
Disclosure > -1.875 s.d.	Despondency = 0	95.7
Disclosure < -1.875 s.d. and > -1.125 s.d.	Despondency = 1	1.4
Disclosure < -2.125 s.d. and > -2.500 s.d.	Despondency = 2	1.9
Disclosure < -2.500 s.d.	Despondency = 3	1.0

Because respondents in most real-life situations are likely to want to present themselves positively in the testing situation, respondents who obtain very low scores on social desirability scales are something of

an anomaly. The reasons for these very low scores may be (a) “faking bad” (b) excessive self-criticism, or (c) despair. The cut-offs for Despondency are set at more extreme values than those for Dissimulation as respondents are more likely to wish a positive than a negative bias to their results.

It is very probable that the pattern of the OBPI scores from respondents who obtain a Despondency score greater than 0 will understate their employment potential, and this must be considered, along with a wider range of background factors, in interpretation. It may also be that the respondent is alienated from their employment environment, adopting a “who cares” attitude. Alternatively, the respondent may be depressed or “crying out for help”.

Inattention

The Inattention audit screen identifies patterns of responding which, although not random, are unrelated to the item content. A case in point may be repeated endorsement of the same response category, or rhythmic alternation of response category endorsement. The raw scale is calculated from the number of changes between consecutive items. Thus, a respondent who answered with only one choice (e.g. only circled “disagree” to all items) would obtain a score of 0 changes, while a respondent who alternated between all items would obtain a raw score of 189 changes. The mean number of changes for the standardisation sample was 114.03, with a standard deviation of 14.46.

The Inattention auditor was derived as follows from these raw numbers of changes.:

Number of changes	Percentage scoring at this value in the standardisation sample	
Between 89 and 141	Inattention = 0	94.4
Between 81 and 88 or between 142 and 149	Inattention = 1	5.2
Between 71 and 80 or between 150 and 154	Inattention = 2	0.4
less than 70 or greater than 154	Inattention = 3	0.0

The Inattention screen is very effective at identifying inappropriate responding of its type.

Scale Development

Item analysis was repeated for the standardisation study using data that had undergone a within-subject standardisation. Some minor adjustments were made to the personality scales in the light of this analysis. Some item replacement was also carried out on the Detail and Emotion scales to further reduce the correlations with socially desirable responding as assessed by the Disclosure scale.

Intercorrelations between the OBPI personality scales are shown in Table 8.

Table 8: Intercorrelations between the OBPI personality scales and correlations with social desirability in the standardisation sample. Social desirability is assessed by the inverse Dissimulation scale, so that a low score on Disclosure is equal to a high score on social desirability and the sign of the correlation coefficient is therefore reversed. Split-half reliabilities are shown in the diagonals.

	F	A	C	E	D	Dis (reversed)
Fellowship	(.73)	.19	-.19	-.21	-.03	.05
Authority		(.77)	-.28	-.25	-.12	-.05
Conformity			(.76)	.15	.28	.11
Emotion				(.81)	-.06	-.20
Detail					(.73)	.29
Disclosure (reversed)						(.76)

Following item analysis, some adjustments were also made to the integrity scales.

To increase reliability, a few of the integrity scales were augmented with items from the personality scales. Where possible, when an item from a personality scale was used to augment an integrity scale, a second item from the same personality scale but which scored in the opposite direction was used as a counter-balance. There is, however, some degree of non-independence between personality and integrity scales, and this will need to be considered in interpretation. When the scores on each integrity scale are being interpreted, some account should be taken of scores on the personality scales with which each integrity scale correlates. This can often illuminate possible reasons for any score patterns found, particularly when relevant biodata information is also considered. The degree of augmentation is given in Table 9

Table 9: Numbers of items from each personality scales used to augment each integrity scales (U = no overlap)

	U	F	A	C	E	D
Proficiency	11	0	0	1	2	0
Work-orientation	15	0	1	1	2	1
Patience	10	0	4	0	0	0
Fair-mindedness	11	1	0	5	2	1
Loyalty	12	0	2	1	1	2
Disclosure	16	0	0	0	0	0
Initiative	12	0	0	1	0	1

Table 10: The intercorrelations between the integrity scales in the standardisation study. (Split-half reliabilities are shown in the diagonal)

	Pro	W-o	Pat	F-m	Loy	Dis	Ini
Proficiency	(.70)	.26	.44	.19	.35	-.53	.11
Work-orientation		(.70)	-.02	.41	-.13	-.27	.47
Patience			(.73)	.20	.42	-.29	-.05
Fair-mindedness				(.72)	-.13	-.08	.44
Loyalty					(.73)	-.32	-.45
Disclosure						(.76)	-.12
Initiative							(.72)

Table 11: Correlations between the personality scales and the integrity scales. (Those marked with a \$ are inflated by augmentation)

	F	A	C	E	D
Proficiency	-.01	-.04	.11\$	-.40\$	.55
Work-orientation	.15	.43\$	-.38\$	-.31\$	.20\$
Patience	-.07	-.43\$	.03	-.22	.29
Fair-mindedness	.25\$	.18	-.62\$	-.32\$	.05\$
Loyalty	-.16	-.52\$	.41\$	.25\$	.52\$
Disclosure	-.05	.05	-.11	.20	-.29
Initiative	.38	.49	-.47\$	-.48	-.07\$

From Table 11 we can see that while both the Proficiency and the Work-orientation scale reflect attitudes to work, high scorers on Proficiency are characterised by high scores on Detail and low scores on Emotion, suggesting that they are both conscientious and able to work under stressful conditions, while higher scorers on Work-orientation tend to obtain high scores on Authority and low scores on Conformity, suggesting that while they are achievement oriented they may tend to be impulsive (see Table 6). High scorers on Proficiency also tend to obtain high scores on Patience and Loyalty, reflecting the distracting effects that a lack of patience or of loyalty may have on proficiency. Work-orientation, on the other hand, correlates with Fair-mindedness and Initiative, the latter suggesting a positive orientation to their working environment. The high negative correlation between Proficiency and Disclosure reflects the fact that a lack of proficiency is normally perceived as particularly socially undesirable. The Disclosure scale can detect a reluctance to disclose socially undesirable attributes. Scores on Disclosure should always be noted carefully when interpreting Proficiency.

Low scorers on Patience tend to obtain high scores on Authority reflecting the common observation that aggression (low Patience) and ambition often go hand in hand. High scorers on Fair-mindedness tend to be high scorers on Fellowship and low scorers on Conformity and Emotion, reflecting both their intellectual propensities and their skills in social leadership (see Table 6). High scorers on Loyalty tend to be low scorers on Authority and high scorers on Conformity, Emotion and Detail, suggesting a willingness to follow orders which may be particularly important in some work environments. The inverse pattern for low scorers on Disclosure suggests a person who is likely, for better or worse, to put their own needs ahead of those of the company (a perfectly acceptable situation so long as these needs coincide). High scorers on Initiative reflect the expected Big Five pattern for entrepreneurs (see Table 6), while Table 10 confirms this pattern, showing while such individuals may be work-oriented and fair-minded, they tend to have low Loyalty scores indicates a preference to be their own boss. Disclosure has relatively low correlations with all five personality scales, confirming the effectiveness of the procedures used to reduce

social desirability correlations for the five personality scales to below 0.3. The correlations of Disclosure with the Proficiency, Work-orientation, Patience and Loyalty integrity scales, on the other hand reflects the fact that low proficiency, low work-orientation, low patience and low loyalty are generally not socially desirable attributes in the working population, and where they exist are less likely to be disclosed.

Correlations with sex, age and educational level

As there were considerable differences in all three of these biodata variables among the various occupations represented in the standardisation study, the pooled within-occupation correlation coefficients, as well as the raw correlations, are shown in Table 12.

Table 12. Correlations of the OBPI variables with age, sex and educational level for the 427 respondents in the standardisation sample.

	Raw correlation			Pooled within occupations		
	Age	Sex	Education	Age	Sex	Education
Fellowship	-.13*	.08	-.02	-.21**	.11	-.04
Authority	.14**	-.06	.17**	.08	-.04	.13*
Conformity	-.08	.05	-.53**	.11	.10	-.34**
Emotion	-.02	.14**	-.02	-.01	.10	-.06
Detail	.04	.07	-.13*	.03	-.02	-.03
Proficiency	.26**	.09	.02	.13*	.02	.01
Work-orientation	.22**	.09	.20**	.09	.02	.07
Patience	.08	.02	.08	.02	-.06	.02
Fair-mindedness	.27**	.01	.53**	.04	-.17**	.35**
Loyalty	.04	.15**	-.17**	.02	.08	-.11
Disclosure	.10*	-.26**	-.10*	-.19**	-.06	.08
Initiative	.07	-.12**	.22**	.01	-.09	.14*

Looking at the Table 12 correlations adjusted for occupation, we see that older respondents obtain lower scores on Fellowship and higher scores on Proficiency. There is also a negative correlation of age with Disclosure, in line with the expectation from the psychological literature that older people have higher social desirability (negative Disclosure) scores. As all these effects are very likely to reflect genuine differences in these traits at various ages no adjustments for age effects are recommended (Rust & Golombok, 2009). While a sex difference is apparently found for Fair-mindedness this should be discounted as no effect existed prior to pooling of within subject-correlations. Educational level correlates significantly with Authority, Conformity, Fair-mindedness and Initiative, all in the expected direction.

Differences between ethnic groups, and between speakers of English as a second language (versus the rest), were examined for each OBPI scale using analysis of variance. No significant differences were found.

VALIDITY

Content validity for the five personality scales of the OBPI

One of the most straightforward ways in which content validity can be reported is to give those items which have the most extreme loadings on each scale in both positive and negative directions. A positive loading means that the item has a large positive correlation with that scale, while a negative loading means that the item has a large negative correlation with that scale. With Fellowship for instance (see below), high Fellowship scorers tend to agree with the item “I am the sort of person who can easily be the life of a party” and to disagree with the item “I hate being the focus of attention”. Low Fellowship scorers, on the other hand, tend to disagree with the item “I am the sort of person who can easily be the life of a party” and to agree with the item “I hate being the focus of attention”. Because of this, the item total correlations of these two items with scores on Fellowship will be positive and negative respectively.

▼ Fellowship

Positive: “I am the sort of person who can easily be the life of a party.”
Negative: “I hate being the focus of attention.”

▼ Authority

Positive: “I don’t care if some people think I am pushy so long as I get things done.”
Negative: “People have sometimes told me I am not forceful enough.”

▼ Conformity

Positive: “Changing the way we do things usually makes matters worse.”
Negative: “I am persistently on the lookout for new ideas to exploit.”

▼ Emotion

Positive: “It is probably true to say that I am something of a worrier.”
Negative: “I would describe myself as being exceptionally free from stress.”

▼ Detail

Positive: “I have a reputation for being good at checking detail carefully.”
Negative: “I find routine administration boring and prefer to leave it to others.”

Content validity of the seven integrity scales of the OBPI

▼ Proficiency

Positive: “Nobody has ever considered me to be unreliable.”
Negative: “On some occasions I find it difficult to concentrate properly on what I am doing.”

▼ Work-orientation

Positive: “My work is more important to me than anything else.”
Negative: “I am sometimes quite happy to leave important decisions to others.”

▼ Patience

Positive: “I am very patient with people even when I know they are wasting my time.”
Negative: “I have sometimes lost my temper with my colleagues.”

▼ Fair-mindedness

Positive: “People are usually honest with me.”
Negative: “Many people are so naive that it’s very easy to manipulate them.”

▼ **Loyalty**

Positive: "I tend to feel uncomfortable if I go against the rules."

Negative: "I like the thrill of taking risks."

▼ **Disclosure**

Positive: "There are times when it is not sensible to tell the truth."

Negative: "I always tell the truth."

▼ **Initiative**

Positive: "I usually find it easy to enthuse others with my ideas."

Negative: "Once some people's minds are made up there is no point in trying to influence them."

Construct validity

The establishment of construct validity for an instrument is a long-term process and may take several decades to develop. This said, however, a strong case can be made for the construct validity of the five OBPI personality scales as they assess the Big Five factors for which the quality of construct validity evidence available in the literature is second to none.

The seven integrity scales of the OBPI are based on a more recent model and new evidence is therefore required. Some is given in the Giotto Handbook which reports a cross validation of the seven integrity scales with a 94-item adjective check list constructed specifically for the Prudentius model (Bergman, 1926). There were 12 adjectives per Prudentius scale in the pilot version of this adjective check list which was administered to 198 staff employed by a personality UK security company. Each set of twelve adjectives was factor analysed to yield a short 8 item interim scale for the 7 Prudentius traits of competence, industriousness, tolerance, equanimity, loyalty, unselfishness and resourcefulness. Each check list scale was scored in such a way that a high score represented a negative (undesirable) attribute. This adjective check list was administered to the respondents in the OBPI standardisation study. The seven checklist scales are detailed below

Checklist scale 1 (Carelessness vs. competence) consisted of 8 adjectives (four positive and 4 negative) the highest factor loadings being for Absentminded, Forgetful, Careless and Thoughtless (negative), and Wise and Serious (positive)

Checklist scale 2 (Absenteeism vs industriousness) consisted of 8 adjectives (four positive and 4 negative) the highest factor loadings being for Irresponsible and Unkempt (negative) and Determined and Tireless (positive)

Checklist scale 3 (Hostility vs tolerance) consisted of 8 adjectives (4 positive and 4 negative), the highest factor loadings being for Stormy, Wild, Aggressive and Angry (negative) and Patient, Accepting and Non-violent (positive).

Checklist scale 4 (Subversion vs. equanimity) consisted of 8 adjectives (4 positive and 4 negative), and the highest factor loadings were for Hard-done-by, Self-righteous and Scapegoated (negative) and Just, Ethical and Impartial (positive).

Checklist scale 5 (Disloyalty vs loyalty) consisted of 10 adjectives (5 positive and 5 negative), and the highest factor loadings were for Opinionated, Egotistical and Arrogant (negative) and Dutiful, Unassuming and Obedient (positive).

Checklist scale 6 (Greed vs. unselfishness) consisted of 8 adjectives (4 positive and 4 negative), and the highest factor loadings were for Unappreciative, Resentful and Envious (negative) and Benevolent, Altruistic and Compassionate (positive).

Checklist scale 7 (Inertia vs resourcefulness) consisted of 12 adjectives (6 positive and 6 negative) and highest factor loadings were for Disheartened, Discouraged and Down-hearted (negative) and Inventive, Creative and Ingenious (positive).

Table 13: The correlations of the seven Prudentius checklist scales with the 12 OBPI scales (N=380)

OBPI scale	Checklist scale						
	1	2	3	4	5	6	7
Fellowship	.04	-.09	.15**	-.05	.18**	-.09	-.23**
Authority	.02	-.33**	.42**	.02	.47**	.04	-.31**
Conformity	-.11*	.27**	-.15*	.34**	-.28**	.22**	.43**
Emotion	.04	.23**	-.01	.14**	-.12*	.08	.46**
Detail	-.43**	-.15**	-.24**	.02	-.38**	-.05	.13*
Proficiency	-.39**	-.24**	-.26**	-.17**	-.29**	-.16**	-.11*
Work-orientation	-.14**	-.48**	.18**	-.19**	.13**	-.14**	-.41**
Patience	-.12*	.06	-.54**	-.21**	-.39**	-.18**	-.03
Fair-mindedness	-.05	-.32**	-.08	-.48**	.05	-.38**	-.39**
Loyalty	-.25**	.12*	-.43**	-.05	-.56**	-.06	.28**
Disclosure	.26**	.17**	.17**	.11*	.23**	.20**	.11*
Initiative	.05	-.32**	.22**	-.20**	.36**	-.17**	-.55**

The correlations between the seven checklist scales and the seven OBPI integrity scales are shown in Table 13. In the correlation matrix, non-significant correlations are notable for their relative paucity. However, some of this surfeit of significance can be attributed to response bias artefacts. The acquiescence effect will not be among these as this has been eliminated through within-subject standardisation for both OBPI and the adjective check list, and by balance of positive with negative items for every scale. However, Prudentius adjectives are more likely to be affected by social desirability. To counterbalance for this influence, the correlation matrix was recalculated as a partial correlation matrix which removed the influence of Disclosure, a robust measure of socially desirable responding. This is shown in Table 14.

Table 14: The partial correlations of the seven Prudentius checklist scales with the 12 OBPI scales, with the social desirability scale (P6: Disclosure) as the partial variate

OBPI scale	Checklist scale						
	1	2	3	4	5	6	7
Fellowship	.06	-.08	.17**	-.05	.20**	-.08	-.23**
Authority	.00	-.37**	.42**	.01	.47**	.03	-.33**
Conformity	-.07	.32**	-.12*	.37**	-.25**	.27**	.46**
Emotion	-.03	.20**	-.06	.11*	-.18**	.03	.45**
Detail	-.37**	-.10*	-.19**	.07	-.33**	.03	.18**
Proficiency	-.37**	-.18**	-.21**	-.13*	-.20**	-.05	-.06
Work-orientation	-.08	-.46**	.24**	-.16**	.20**	-.10	-.39**
Patience	-.04	.12*	-.52**	-.19**	-.34**	-.12*	.00
Fair-mindedness	-.03	-.31**	-.07	-.47**	.07	-.38**	-.38**
Loyalty	-.17**	.20**	-.40**	-.02	-.52**	.01	.34**
Initiative	.08	-.31**	.24**	-.19**	.40**	-.15**	-.54**

It can be seen in Table 14 that removing the effect of social desirability eliminates some of the smaller significant correlations but leaves most unchanged. The pattern of intercorrelations found for the Big Five conforms to expectation. They generally confirm the interpretations given in the literature for Big Five profile scores.

A more robust requirement for validation of the seven integrity scales can now be included. It will be remembered that one of the claims for the Big Five model is that it is a complete description of personality. If this was the case, then we would not expect the integrity scales to add anything to the total picture over and above the information already incumbent in the Big Five. We can evaluate this claim by calculating partial-correlation coefficients that remove the effect of all five personality scales from the correlation matrix between the OBPI and the Prudentius checklist scales. This partial correlation matrix is shown in Table 15.

Table 15: The partial correlations of the seven Prudentius checklist scales with the 12 OBPI scales, with the social desirability scale (Disclosure) and the five personality scales as partial variates

OBPI scale	Checklist scale						
	1	2	3	4	5	6	7
Proficiency	-.16**	-.13*	-.13*	-.17**	-.10	-.07	-.03
Work-orientation	-.01	-.22**	.10	-.07	-.02	-.03	-.19**
Patience	.02	.06	-.40**	-.16**	-.15**	-.09	-.03
Fair-mindedness	-.03	-.10	-.19**	-.33**	-.12*	-.28**	-.04
Loyalty	-.01	.01	-.20**	-.18	-.23**	-.05	-.03
Initiative	.06	-.04	.05	-.05	.15**	-.08	-.23**

Based on this analysis the following further evidence for the validity of the integrity scales is suggested:

- ▼ **Proficiency.** The correlation with adjectival competence (negative Carelessness) remains significant. This suggests that the Proficiency scale is valid as a predictor of competence and provides data on this trait over and above that already implicit in the Detail scale.
- ▼ **Work-orientation.** The negative correlation with industriousness (negative Absenteeism) supports the validity of P2 as a work-orientation scale.
- ▼ **Patience.** The negative and very significant correlation of Patience with tolerance (negative Hostility) provides very strong evidence for the validity of this scale as a measure of overt aggression. This is particularly the case as covert aggression from the Authority scale has been eliminated in this matrix.
- ▼ **Fair-mindedness.** The correlation of Fair-mindedness with equanimity (negative Subversion) is extremely encouraging.
- ▼ **Loyalty.** Evidence for the validity of this scale is provided by the significant correlation with the checklist scale of the same name (negative Disloyalty).
- ▼ **Initiative.** The significant correlation with resourcefulness (negative Inertia) is consistent with many of the negative attributes of Initiative.

Criterion-related validity

The five OBPI Five Personality Scales were validated against 10 predesignated supervisors rating scales (one positive and one negative rating for each scale). Ratings ranged from below average through average, a little above average, much above average to exceptional. The ratings were carried out by the supervisors of 214 respondents in the standardisation study. For each supervisor, their responses to the rating items were standardised within the response set, so that appraisal ratings were in comparison with their average rating rather than absolute. There was some missing data, so the N for the correlations given below ranges between 191 and 214. The results are shown in Table 16.

Table 16: Correlations of the five OBPI personality scales with supervisors' ratings (N=200 approximately)

	Fellowship
Team skills	.31**
Ability to work independently	-.23*
	Authority
Ability to make tough decisions	.39**
Ability to make friends with colleagues	-.21+
	Conformity
Obedience to company policy	.37**
Ability to generate new ideas	-.41**
	Emotion
Tendency to worry	.13
Level of self-confidence	-.39**
	Detail
Attention to detail	.38**
Breadth of vision	-.36**

Validity coefficients adjusted for attenuation (assumed reliability of supervisor's rating = 0.50).
Significance levels (based on unadjusted correlations):

**** = $p < .01$ (two-tailed), * = $p < .05$ (two-tailed). + = $p < .05$ (one-tailed)**

Because of the nature of the integrity scales it is not appropriate to pool the supervisors' ratings for all the respondents in the validation study, as what counts as 'appropriate behaviour' insofar as these scales are concerned varies considerably from occupation to occupation. However, a correlational analysis relating supervisors' ratings to scores on the integrity scales was carried out for the largest pool of subjects within the standardisation study which consisted of 61 security escort personnel. The results of this are shown in Table 17, and, notwithstanding the small sample size, are broadly supportive of the criterion-related validity of the seven OBPI integrity scales.

Table 17: Correlations of the seven OBPI integrity scales with supervisors' ratings for 61 security escort personnel. Supervisors rated employees as either below average, average, above average, much above average or exceptional of these traits. Ratings for each of these categories were converted to the numbers 1 to 5 respectively.

	Proficiency
Time-keeping	.56*+
Performance under pressure	.46*
	Work-orientation
Attention to detail	.47*
	Patience
Level of self-control	.43*
Dependability	.45*
	Fair-mindedness
Trustworthiness	.48*
Ability to work to strict guidelines	.43*
	Loyalty
Obedience to company policy	.43*
Ability to make friends with colleagues	.43*
	Disclosure
Likelihood of making mistakes	-.44*
	Initiative
Level of self-confidence	.47*

Validity coefficients are adjusted for attenuation (assumed reliability of supervisor's rating = 0.50), Significance levels are based on unadjusted correlations:

*** = $p < .05$, *+ = $p < .02$ (both two-tailed)**

REFERENCES

- Allport, G. W. & Odbert, H. S. (1936) Trait-names: A psycholexical study. *Psychological Monographs*, 47, (1 While No. 211)
- Almagor M, Tellegen A & Waller N G (1995) The Big Seven model: A cross-cultural replication and further exploration of the basic dimensions of natural language trait descriptors. *Journal of Personality and Social Psychology*, 69, 300-307.
- Anastasi & Urbina (1998) *Psychological Testing* (7th Edition) Macmillan, New York
- Barrick M R & Mount M K (1991) The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, No. 1, 1-26.
- Bass, B. H. (1955) Authoritarianism or acquiescence? *Journal of Abnormal and Social Psychology*, 51, 616-623.
- Bergman, J (1926) *Aurelii Prudentii Clementis. Corpus Scriptorum Ecclesiasticorum Latinorum*, 61, Hoelder-Pichler-Tempsky A.G.: Vienna.
- Camara W J & Schneider D L (1994) Integrity tests: Facts and unresolved issues. *American Psychologist*, 49, 112-119
- Cattell, R. B. (1943) The description of personality: basic traits resolved into clusters. *Journal of Abnormal and Social psychology*, 38, 476-506.
- Cattell, R. B. (1995) The fallacy of the five factors of personality. *The Psychologist*, May 207-208.
- Claridge, G. S. & Birchall, P. M. (1978) Bishop, Eysenck, Block and psychoticism. *Journal of Abnormal Psychology*, 87, 604-668.
- Collins J M & Schmidt F L (1993) Integrity, personality and white-collar crime: A construct validity study. *Personnel Psychology*, 46, 295-311.
- Conoley, J. J. & Kramer, J. C. (1989) (Eds) *Tenth Mental Measurement Yearbook*. Buros Institute of Mental Measurement. University of Nebraska Press, Lincoln NE
- Costa, P. T. & McCrae, R. R. (1992) Four ways five factors are basic. *Personality and Individual Differences*, 13, 653-665.
- DeYoung, C. G., Carey, B. E., Krueger, R. F., & Ross, S. R. (2016). Ten aspects of the Big Five in the Personality Inventory for DSM-5. *Personality Disorders: Theory, Research, and Treatment*, 7(2), 113-123.
- Digman, J. M. (1990) Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- Eysenck, H. J. (1947) *Dimensions of Personality*. Routledge & Kegan Paul, London.
- Fiske, D. W. (1949) Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, 44, 329-344

- Galton F. (1884) Measurement of character. *Fortnightly Review*, 42.
- Goldberg, L. R. (1982) From Ace to Zombie: Some explorations in the language of personality. In CD Spielberger & JN Butcher (Eds) *Advances in personality Assessment*. Lawrence Earlbaum, Hillsdale NJ.
- Goldberg (1990) An alternative 'Description of Personality'. The big five structure. *Journal of personality and social psychology*, 59, 1216-1229
- Goldberg L R, Grenier J R, Guion R M, Sechrest L B & Wing H (1991) Questionnaires used in the prediction of trustworthiness in pre-employment selection decisions. Washington DC; American Psychological Association
- Goldberg, L. R. (1992) The development of markers of the big five structure, *Psychological Assessment*, 4, 26-42.
- Gough H G (1971) The assessment of wayward impulse by means of the Personal Reaction Blank. *Personnel Psychology*, 24, 669-677.
- Guastello S J & Rieke M L (1991) A review and critique of honesty test research. *Behavioral Sciences and the Law*, 9, 501-523.
- Hogan R (1986) *Hogan Personality Inventory Manual*. National Computer Systems; Minneapolis. , 153-159.
- John, O. P. (1990) The big five factor taxonomy: Dimensions of personality in natural language and in questionnaires. In L. A. Pervin (Ed) *Handbook of Personality Theory and Research*. Guilford; New York.
- Klump C S (1964) *Stanton Survey*, Stanton Corporation; Charlotte NC.
- Knowles, J. B. (1963) Acquiescence response set and questionnaire measurement in personality. *British Journal of Social and Clinical Psychology*, 2, 131-137.
- Legislation.gov.uk (2010). Equality Act 2010. (online) Available at: <http://www.legislation.gov.uk/ukpga/2010/15/contents>.
- Legislation.gov.uk (2018) Data Protection Act 2018 (online). Available at <http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>.
- Lilienfeld, S. O., Alliger, G. & Michell, K. (1995) Why integrity testing remains controversial. *American psychologist*, 50, 457-458.
- Loehlin, J. (1992) *Latent variable models*. Earlbaum; Hillsdale NJ.
- Loevinger J (1994) Has psychology lost its conscience? *Journal of Personality Assessment*, 62, 2-8.
- London House Press (1980) *Personnel Selection Inventory Manual*. Author; Park Ridge, IL.
- London House Press (1982) *Employee Attitude Survey*. Author; Park Ridge IL. Lousig-
- Nont G M (1987) *Phase II Profile*. Lousig-Nont and Associates; Las Vegas NV.

Mershon, B. & Gorsuch, R. L. (1988) Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology*, 55, 675-680.

Miller J F & Bradley P (1975) *Milby Profile*. Milby Systems; Minneapolis MN.

Norman, W. T. (1967) 2800 personality trait descriptors: Normative operating characteristics for a university population. Ann Arbor, University of Michigan, Department of Psychology.

O'Bannon R M, Goldinger L A & Appleby G S (1989) *Honesty and integrity testing*. Atlanta GA: Applied Information Resources.

Ones D S, Viswesvaran C & Schmidt F L (1993) Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. *Journal of Applied Psychology*, 78, 679-703.

Ones D S Viswesvaran C & Schmidt F L (1995) Integrity tests: Overlooked facts, resolved issues, and remaining questions about predictive validity, construct validity, and in-house research. *American Psychologist*, 50, 456-457.

Ones D S, Schmidt F L, Viswesvaran C & Lykken D T (1996) Controversies over integrity testing: Two viewpoints. *Journal of Business and Psychology*, 10, 487-501.

Paajanen G E (1987) The prediction of counterproductive behavior by individual and organisational variables. Unpublished Doctoral Dissertation; University of Minnesota, Minneapolis.

Paulhus, D. L. (1984) Two-component model of socially desirable responding. *Journal of Personality and Social Psychology*, 4, 44-49.

Paulhus, D. L. (1991) Measurement and control of response bias. In P J Robinson, Shaver & Wrightsman (Eds) *Measures of Personality and Social Psychological Attitudes*, Academic Press, NY.

Peabody, D. (1987) Selecting representative trait adjectives. *Journal of Personality and Social Psychology*, 52, 59-71.

Peabody, D. & Goldberg, L. R. (1989) Some determinants of factor structures from personality trait descriptors. *Journal of Personality and Social Psychology*, 57, 552-567.

Reid Psychological Systems (1951) *Reid Report*. Author; Chicago

Rust, J. (1975) Genetic effects in the cortical evoked potential. *Electroencephalography and Clinical Neurophysiology*, 39, 321-327.

Rust, J. (1975) Cortical evoked potential, personality and intelligence. *Journal of Comparative and Physiological Psychology*, 89, 1220-1226.

Rust, J. (1987) The Rust Inventory of Schizotypal Cognitions: A psychometric measure of psychoticism in the normal population. *British Journal of Clinical Psychology*, 26, 151-152.

Rust, J. (1988) Implications of connectionism for psychophysiology. *Journal of Psychophysiology*, 2, 87-89.

- Rust, J. (1988) The Rust Inventory of Schizotypal Cognitions (RISC). *Schizophrenia Bulletin*, 14, 317-322.
- Rust, J. (1989) Manual of the Rust Inventory of Schizotypal Cognitions. The Psychological Corporation; London.
- Rust, J. (1988) Socio-biology and psychometrics: Do they really need each other? *Philosophical Psychology*, 1, 117-129.
- Rust, J. (1996) Giotto Handbook, The Psychological Corporation, London
- Rust, J. (1996) Orpheus Handbook, The Psychological Corporation, London
- Rust, J. & Golombok, S. (2009) Modern Psychometrics, 3rd Edition, Routledge, London
- Sackett P R, Burris L R & Callahan C (1989) Integrity testing for personnel selection: An update. *Personnel Psychology*, 37, 491-529.
- Sackett, P. R. & Wanek, J. E. (1996) New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness and reliability for personnel selection. *Personnel Psychology*, 49(4), 787-830.
- Sayers, D (1955) Dante's Purgatorio (English translation). Penguin books.
- Schmidt F L, Ones D S & Hunter J E (1992) Personnel selection. *Annual Review of Psychology*, 43, 627-670.
- Schmidt J M & Ryan A (1993) The big five in personnel selection: Factor structure in applicant and non-applicant populations. *Journal of Applied Psychology*, 78, 966-974.
- Science Research Associates (1987) Personal Outlook Inventory (POI) Examiner's Manual. Author; Chicago.
- Tett, R. P., Jackson, D. N. & Rothstein (1991) Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, No. 4, 703-742.
- Thompson, H. J. (1949) Prudentius (Latin with an English translation). Harvard University Press: Cam MA.
- Thurston, L. L. (1934) The vectors of the mind. *Psychological Review*, 41, 1-32.
- Tupes, E. C. & Crystal, R. E. (1961) Stability of personality factors based on trait ratings (USAF ASD Technical report No. 61-97). Lackland Air Force Base, TX US Air Force
- US Congress, Office of Technology Assessment (1990) The use of integrity tests for pre-employment screening (OTA-SET-442). US Government Printing Office, Washington DC.