


ORPHEUS MANUAL

JOHN RUST

 The Psychological Corporation®
A Harcourt Assessment Company

Published by The Psychological Corporation Limited,
Harcourt Place, 32 Jamestown Road, London NW1 7BY.
Copyright © 2001, 1996 by The Psychological
Corporation.

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

Printed in the United Kingdom. ISBN 0 7491 0660 3

A B C D E
1 2 3 4 5

CONTENTS

PART 1: ORPHEUS USER MANUAL	1
Brief description of Orpheus	2
Purpose	3
The Orpheus scales	4
The five major scales	4
The seven minor scales	5
Advantages of Orpheus	6
Instructions for use	7
Before testing begins	7
The testing environment	7
Establishing and maintaining rapport	8
Administering Orpheus	9
Administration by computer	10
Keeping records	11
Scoring Orpheus	12
Missing data	12
The scoring software	12
Output from the scoring software	13
Response Audit	13
Standardisation	15
Reliability and validity	17
Reliability	17
Validity	18
Interpretation of Orpheus	19

CONTENTS

PART 2: ORPHEUS TECHNICAL MANUAL	21
Background to Orpheus	22
Source of items	22
Test specification	23
The five Orpheus major scales	23
Stability of the “Big Five” model of personality	25
The seven Orpheus minor scales	33
Integrity testing	33
The Prudentius model	38
The construction of the Orpheus scales	40
The major scales	40
The minor scales	43
The pilot study	45
Description of the sample	45
Item analysis	45
The standardisation study	46
Description of the sample	46
Statistical analysis	46
Development of the Response Audit scales	46
Scale development	50
Correlations with sex, age and educational level	55
Reliability	58
Validity	59
Content validity of the five major scales of Orpheus	59
Content validity of the seven minor scales of Orpheus	60
Construct validity	60
Criterion-related validity	65
REFERENCES	69

TABLES

Table 1:	The frequency with which each Response Audit score was obtained in the standardisation sample	14
Table 2:	Orpheus scale stanine to percentile equivalents assuming a normal distribution	16
Table 3:	Split-half reliability and standard error of measurement obtained from the standardisation sample for the five Orpheus major scales and the seven Orpheus minor scales	17
Table 4:	Domains and trait specification for the five Orpheus major scales	25
Table 5:	Names given to “Big Five” traits in the literature	26
Table 6:	Some interpretations of specific five factor profiles that have appeared in the literature	31
Table 7:	The classical theory of the Psychomachia of Prudentius together with the matching integrity traits from Giotto and the seven Orpheus minor scales	39
Table 8:	Intercorrelations between the Orpheus major scales and correlations with social desirability in the standardisation sample	51
Table 9:	Numbers of items from each of the major scales used to augment each of the minor scales	52
Table 10:	The intercorrelations between the minor scales in the standardisation study	53
Table 11:	Correlations between the major scales and the minor scales	54
Table 12:	Correlations of Orpheus variables with age, sex and educational level for the 427 respondents in the standardisation sample	56
Table 13:	The correlations of the seven Prudentius checklist scales with the twelve Orpheus scales	62

CONTENTS

Table 14:	The partial correlations of the seven Prudentius checklist scales with the twelve Orpheus scales, with the social desirability scale (P6: Disclosure) as the partial variate	63	—
Table 15:	The partial correlations of the seven Prudentius checklist scales with the seven Orpheus minor scales, with the social desirability scale (P6: Disclosure) and the five major scales as partial variates	64	
Table 16:	Correlations of the five Orpheus major scales with supervisors ratings	66	
Table 17:	Correlations of the seven Orpheus minor scales with supervisors ratings for 61 security escort personnel	67	—

ORPHEUS MANUAL

PART 1: ORPHEUS USER MANUAL

BRIEF DESCRIPTION OF ORPHEUS

BRIEF DESCRIPTION OF ORPHEUS

Orpheus is a broad spectrum work-based personality questionnaire. It contains 190 items, and generates scores on sixteen scales - five major scales, seven minor scales, and four audit scales. The major scales are Fellowship, Authority, Conformity, Emotion and Detail and are based on the "Big Five" model of personality. The minor scales are Proficiency, Work-orientation, Patience, Fair-mindedness, Loyalty, Disclosure and Initiative, and are based on the Prudentius model of integrity. The four response audits are Dissimulation, Ambivalence, Despondency and Inattention, and are designed to screen for inappropriate responding. Shorthand codes are F, A, C, E and D for the major scales, P1 to P7 for the minor scales, and RA1 to RA4 for the response audits. The Orpheus package consists of (a) a printed question and answer booklet; (b) user and technical manual; and (c) software. The software, as well as scoring Orpheus and administering the test when required, prints a full technical report and a personalised narrative report.

PURPOSE

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

THE ORPHEUS SCALES

The five major scales

The five Orpheus major scales are based on the “Big Five” model of personality, interpreted within the context of work-related behaviours, beliefs, attitudes and interests. The five Orpheus major scales (Fellowship, Authority, Conformity, Emotion and Detail) represent social, organisational, intellectual, emotional and perceptual aspects of personality respectively. The major scales are largely independent of one another and therefore they can additionally be interpreted in combination so that particular meanings can be assigned to certain patterns of scores. For example, a person with high Fellowship and high Authority (assertiveness) can be meaningfully differentiated from a person with low Fellowship and low Authority (submissiveness). Where appropriate, these combinations appear in the narrative report.

F: Fellowship assesses the “Big Five” trait of extraversion/introversion. High F scorers are generally happier working with others or in a team. Low F scorers generally prefer work that requires a degree of independence.

A: Authority assesses the “Big Five” trait of tough vs. tender-mindedness, sometimes called “agreeableness”. High A scorers can make tough decisions. Low A scorers generally adopt a more co-operative approach.

C: Conformity assesses the “Big Five” trait of “openness-to-experience”. High C scorers are likely to have a preference for traditional ways of doing things and to respect established values. Low C scorers often wish to do things differently, and to seek out alternative solutions to problems.

E: Emotion assesses the “Big Five” trait of neuroticism. High E scorers, while often being of a nervous disposition, are likely to be sensitive to the feelings of others. Low E scorers are likely to be more able to perform under stressful conditions but may lack caution.

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

The seven minor scales

The Seven minor scales of Orpheus 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 particular characteristics or actions for others. A weakness in one occupation may be neutral, or even a strength in another and therefore the minor scales should only be used where relevant to a particular 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.

P1: 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.

P2: 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.

P3: Patience assesses the ability to control aggression in whatever form. It is of relevance to work environments where bullying has been a particular concern.

P4: Fair-mindedness assesses fairness in judging the actions of others. It is of relevance to work environments that are beset with strife.

P5: 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.

P6: 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.

P7: 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 ORPHEUS

- Produces a personality profile based on the “Big Five” model of personality within a work context.
- Produces for the first time seven additional personality 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, acquiescence and gender.
- 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 either paper and pencil form or by computer.
- Incorporates state-of-the-art psychometric procedures in scoring, including within-subject standardisation and adjustment for missing data.
- Quickly and easily scored by computer to give full technical and narrative 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 the Data Protection Act (1984), the Sex Discrimination Act (1975), the Race Relations Act (1976) or the Disabilities Discrimination Act (1995).

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 written 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. This information should always include full name, date of birth, sex and current employment, and will usually also include some details of employment and occupational history. Information on race may be requested in order to enable ethnic monitoring to take place.

The testing environment

Orpheus 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, conduct the testing 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.

INSTRUCTIONS FOR USE

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. The administrator should explain the written 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. These are (a) leaving items out, (b) circling two responses for an item (e.g. both "agree" and "disagree") and (c) changing the question before answering. If rapport is properly established and maintained, however, sabotage is much less likely. Remember that Orpheus cannot be scored if more than S items have been left out or sabotaged. If this happens the administrator's, candidate's and organisation's time have been wasted. Always check completed questionnaires to ensure that all questions have been answered.

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 his or her 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 lack of understanding is 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 Orpheus 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 Orpheus was not standardised with such modifications. For example, if sign language or reading aloud 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 Orpheus

Candidates should be told what will happen during the testing session, how long it is likely to take, and whether or not they may leave as soon as they have completed the test. They may also wish to know whether or not they are allowed to smoke. Orpheus 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, smoking and early leaving can be disruptive to other candidates. Also, it is important to be able to check completed questionnaires for missing data. It is relatively easy to point out a few missed items and encourage the candidate to have another go. Once they are out of the building, however, this will normally be impossible and everyone's time may have been wasted. Remember, personality questionnaires can only be scored if candidates have answered the questions (unlike knowledge-based tests in which missed items are simply counted as "wrong").

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

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 his or her scores.

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

INSTRUCTIONS FOR USE

Administration by computer

If Orpheus is to be administered by computer, then particular care must be taken to ensure that the candidates understand what is expected of them. An opportunity to ask questions should be given after the candidate has read the instructions and carried out the trial items. The administrator should always be available during the session to answer any questions or queries the candidate may have.

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 marked on all forms and completed questionnaires. Each completed form should be checked for clerical errors.

Completed questionnaires should be kept in a locked store and should be accessed only by those with authority to do so. In no circumstances should completed questionnaires be left lying around an open office. Disposal of completed questionnaires should be by shredding and/or incineration.

SCORING ORPHEUS

Orpheus can only be scored by software. There is no provision for hand scoring. This is because of the complex mathematical procedures that are involved in the scoring process. These include within-subject standardisation of responses, Response Audit, estimation of missing data, and standardisation to norm groups.

Missing data

Orpheus uses its own internal procedure for dealing with missing data, although this will only operate if the amount of missing data is so small as to not risk affecting the reliability of the scores. If more than five items are missed out, the data set is rejected. If five items or less are missed out, then the missing items are replaced by the mean scores for those items from the standardisation sample. An item counts as missing if (i) no response is made (ii) more than one response is circled (e.g. both agree and disagree) or (iii) the question has been changed and a response endorsed in such a manner as to suggest that the response is to the altered question.

The scoring software

The Orpheus software operates under Windows 3.1. It requires a minimum of about 4mb of core memory, but is much quicker with 6mb or more. Hard disk space of about 5mb is required. It operates under any processor speed, but runs more quickly with faster processors. The software is supplied on three floppy disks. To install the software place disk 1 in the A drive, open the "File" menu of the Program Manager window, select "run", then type ":a:\setup". Follow the instructions that then appear on the screen, and the software will install itself on your computer.

Security of the software is maintained with the use of a password. You should change this from the default "EURYDICE" as soon as your software is installed. The software also requires a dongle which is supplied as part of the start-up kit and can be simply fixed to your machine's parallel port (the printer can still operate as this is plugged into the other socket of the dongle). This dongle is initially loaded with a number of credits which are then used for the computer administration of Orpheus as well as the generation of technical and narrative reports. Further credits can be obtained from The Psychological Corporation at any time.

The Orpheus software is initiated by clicking on the Orpheus icon in Microsoft Windows. The software has been designed to be user friendly, and immune from meddling. If you press the wrong key it will tell you. The operation should be self evident, but, if not, use trial and error and it should soon be picked up. Up-to-date

details of any changes or important points concerning the software's use are provided on the "readme" file which can be found in the c:\orpheus directory once the software has been installed, and also on Installation Disk 3.

Output from the scoring software

There are two forms of output (a) a full technical report, that includes details of all scores obtained by the respondent, and (b) a narrative report. Interpretation of both of these should only be carried out by properly qualified personnel who are also familiar with the technical manual.

The technical report contains three sections (a) the Response Audit (b) the major scale scores and (c) the minor scale scores.

Response Audit

Orpheus conducts a Response Audit at the time of scoring. This assigns a value of 0, 1, 2 or 3 to each of the four audit measures. PA 1 (Dissimulation) is a social desirability score, based on extreme scores on P6: Disclosure, RA2 (Ambivalence) assesses contradiction, and is calculated from a set of similar and opposite items within the questionnaire, RA3 (Despondency) assesses faking bad (relevant to some testing situations, e.g. injury insurance claims), and RA4 (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.

If the sum of the scores of these four measures is 3 or greater, the scoring software will still print the technical report but will include a warning that the results are of questionable value. No narrative report will be produced.

SCORING ORPHEUS

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

	0	% obtaining this score		
		1	2	3
RA1: Dissimulation	88.9	10.1	7.5	2.9
RA2: Ambivalence	89.3	10.7	2.7	0.2
RA3: Despondency	95.7	4.3	2.9	1.0
RA4: Inattention	94.4	5.6	0.4	0.0

STANDARDISATION

Both major scale and minor 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 on the basis of its relationship to the raw scores of a representative group of respondents, referred to as the norm group. Details of the norm group for Orpheus are presented in the Technical Manual in Part 2 of this Manual. The particular form of standardisation used by Orpheus is called the 'stanine'. The stanine scores on a norm group have a mean of 5 and a standard deviation of 2 by definition. If any individual respondent from the norm group has a stanine score of 5, this means that he or she is at the average for that group on the scale in question. His or her 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 he or she is 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 particular scores can be obtained from z tables in any good statistics book).

The stanine transformation is carried out using the following equation

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

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.

STANDARDISATION

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

<i>Stanine score</i>	<i>s.d. range</i>	<i>Percentile equivalents</i>
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 and are obtained from z tables. Actual frequencies may differ.

RELIABILITY AND VALIDITY

Reliability

The split-half reliabilities for the 12 Orpheus scales from the 423 respondents in the standardisation sample are given in Table 3.

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

	Reliability	Standard error of measurement
F: Fellowship	.73	1.04
A: Authority	.77	0.96
C: Conformity	.76	0.98
E: Emotion	.81	0.87
D: Detail	.73	1.04
P1: Proficiency	.70	1.10
P2: Work-orientation	.70	1.10
P3: Patience	.73	1.04
P4: Fair-mindedness	.72	1.06
P5: Loyalty	.73	1.04
P6: Disclosure	.76	0.98
P7: Initiative	.72	1.06

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

RELIABILITY AND VALIDITY

Validity

For details of the validation of Orpheus you are referred to the Technical Manual in Part 2 of this Manual.

INTERPRETATION OF ORPHEUS

The Orpheus software produces a technical report which summarises the scores on the twelve scales and the four Response Audits. This report should only be interpreted by an appropriately trained professional who is familiar with the details provided in the Technical Manual that appears in Part 2 of this Manual. The software also provides a personalised narrative report based on the five major scales individually and in combination.

—

—

—

—

ORPHEUS MANUAL

PART 2: ORPHEUS TECHNICAL MANUAL

BACKGROUND TO ORPHEUS

Source of items

Orpheus 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

Orpheus is a purpose-led personality test which aims to cover a wide scope of potential applications in respect to personality test usage in employment settings. Two separate approaches to trait targeting are used. Firstly, Orpheus has been specifically designed to produce a “Big Five” personality profile. Secondly, Orpheus 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 on the basis of their wide applicability. For each target scale approximately equal numbers of positive and negative items were included in the pilot version.

The five Orpheus major 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 particular five trait 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 Schmit & 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 (Norman, 1963, 1967; Norman & Goldberg, 1966; Passini & Norman, 1966) 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

TEST SPECIFICATION

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 particular 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 of the descriptors and had therefore had to rely on smaller subsets. Norman's initial very extensive work was followed up by Goldberg (1981, 1982) who factor analysed 1,710 of Norman's trait adjectives. In a number of 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 (Digman, 1989, 1990; Digman & Inouye, 1986) and by John (1989, 1990) also supported the fundamental role of the “Big Five” model.

Costa and McCrae (1992) argue that the “Big Five” 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) (Tupes & Cristal, 1958; Church & Katigbak, 1989; Kohnstamm et al, 1995; McCrae, 1989; McCrae & Costa, 1987; Ostendorf & Angleinter, 1994; Saucier, 1992) (ii) they are compatible with a wide variety of psychological theories including psychoanalytic, psychometric and folk-psychological (Costa et al, 1986; Costa & McCrae, 1989; Costa, 1991; McCrae, 1992; McCrae et al, 1995; Miller, 1991; Widiger & Trull, 1992) (iii) they occur in many different cultures (Angleinter et al, 1990; Caprara et al, 1994; DeRaad & Szirmak, 1994; Sztrmak & De Raad, 1994; Kashiwagi et al, 1994; Pulver et al, 1995; Yang & Bond, 1990) 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 Orpheus, the “Big Five” model is re-conceptualised as a domain theory of personality, i.e. each of the “Big Five” factors is considered to be unique to a particular 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 particular person's position with respect to each of these five domains we have an almost complete description of the functioning of his or her personality.

The Orpheus 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 Orpheus major scales

<i>Domain</i>	<i>Orpheus trait</i>	<i>"Big Five" trait</i>
Social	Fellowship	Extraversion vs. Introversion
Organisational	Authority	Tough-mindedness vs. Agreeableness,
Intellectual	Conformity	Conventional vs. Openness-to-experience
Emotional	Emotion	Neuroticism vs. Confidence
Perceptual	Detail	Conscientiousness

Stability of the "Big Five" model of personality

While considerable evidence has accumulated to show that the five factor solution is more stable than that any other number 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 particular 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. The source is given alongside the trait name in brackets

Social domain

Extraversion (Eysenck, 1947, 1992; Briggs 1992)
 Surgency, Talkative, Assertive, Energetic (Tupes and Christal, 1961)
 Impulsiveness, Sociability (Eysenck, 1970)
versus
 Introversion (Eysenck 1947; Myers, 1985; McCrae, 1989)

Organisational domain

Aggression (Murray, 1938)
 Tough-mindedness (James, 1907)
 Moving-against tendency (Homey, 1945)
 Inhibition (Strelau, 1990)
 Narcissism (Freud, 1957)
 Hostility (Dembroski, 1988)
 Coronary proneness (Costa, 1987)
 Indifference-to-others, Self-centredness (Digman, 1990)
 Spitefulness, jealousy (Digman, 1990)
 Authoritarianism (Adorno, 1950)
 Hostile-non-compliance (Digman and Takemoto-Chock 1981)
 Antagonism (Graziano and Eisenberg, 1993)
 Type A behaviour (Costa, 1989)
 Thinking (Myers, 1985; McCrae, 1989)
versus
 Abasement, Nurturance (Murray, 1938)
 Trust (Erikson, 1950)
 Self-monitoring (Synder, 1974)
 Social interest (Adler, 1964)
 Good-natured, Co-operative, Trustful (Tupés and Chnstal, 1961)
 Altruism, Nurturance, Caring, Emotional-support (Digman, 1990)
 Friendly-compliance (Digman and Takemoto-Chock, 1981)
 Friendliness (Guilford and Zimmerman, 1949)
 Tender-mindedness (James, 1907; Avia, 995)
 Feeling (Myers, 1985; McCrae, 1989)

Intellectual domain

Dogmatism (Rokeach, 1954, 1960)
 Sensing (Myers, 1985; McCrae, 1989)
versus

Openness-to-experience, creativity, divergent thinking (McCrae, 1987, 1993)
 Understanding, Change, Sentience, Autonomy (Murray, 1938)
 Experience-seeking (Zuckerman, 1993)
 Absorption (Tellegen, 1985)
 Flexibility, Achievement-via-independence (Gough, 1987)
 Artistic interests (Holland, 1985)
 Private Self-consciousness (Buss, 1989)
 Culture (Norman, 1963)
 Intellectual, Cultured, Polished (Tupes and Christal, 1961)
 Differentiated emotions, Aesthetic sensitivity, Need-for-variety,
 Unconventional values (McCrae and John, 1992)
 Intuition (Myers, 1985; McCrae, 1989)

Emotional domain

Anxiety (Taylor, 1948)
 Neuroticism (Eysenck, 1985)
versus
 Emotional stability, Calm, Not neurotic, Not easily upset (Tupes and Christal, 1961)

Perceptual domain

Achievement, Order, Endurance (Murray, 1938)
 Persistence (Windel and Lernem, 1986)
 Competence (White, 1959)
 Constraint (Tellegen, 1982)
 Prudence (Hogan, 1986)
 Will-to-achieve (Digman and Takemoto-Chock, 1981)
 Superego strength (Cattell, 1970)
 Locus of control (Rotter, 1966)
 Character (Hartshorn, 1929)
 Persistence (Lorr, 1986)
 Achievement-motive (McClelland, 1953; Atkinson, 1977)
 Dependability, Conscientiousness, Responsibility, Orderly (Tupes and Christal, 1961)
 Perceiving (Myers, 1985; McCrae, 1989)
versus
 Judging (Myers, 1985; McCrae, 1989)

Note: In many cases the link of the trait name with the "Big Five" has been suggested by an author other than the originator.

TEST SPECIFICATION

In spite of the consensus that five factors are optimal, for many of the factors there is rather less agreement about what the constructs mean. Why are some of these factors so difficult to tie down? There are several reasons for this. The first arises from the grounding of the “Big Five model in the statistical procedure of factor analysis. In spite of 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, 1994), 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 make 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 particular issue which has received special attention in the construction of Orpheus 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, 1990; Furnham, 1986, 1990; Furnham & Henderson, 1982; Wiggins, 1966). 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 (Edwards, 1957; Crowne & Marlowe, 1960; 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; Couch & Keniston, 1960). This effect can also effect 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 num-

TEST SPECIFICATION

ber 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 (Schmit & 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 themselves 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 (Vassend & Skrondal, 1995). It describes the most frequent solution under the majority of circumstances. The particular 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 con-

joined. 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 can be attained, the five factor model can provide more information than a questionnaire which claims to assess 70 oblique factors.

Table 6: *Some 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
Impulsivity	High Authority, Low Conformity
Authoritarian	High Authority, High Conformity

TEST SPECIFICATION

Table 6 shows some common interpretations that have been made for particular “Big Five” profiles. It can be seen that 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 particular 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 Orpheus itself— this is the question of whether the latent traits of the “Big Five” could represent real entities which may actually 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 currently so much in 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 particular degree of relationship over any other. To put this another way, if we are no longer specifically targeting a correlation of 0.0 between the factors, should we target a between-factors correlation of 0.1, or 0.2, or 0.5 or 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 psy-

chophysiology provides a mediator between personality and biology (Rust, 1976, 1977) there is no necessary link between the two (Rust, 1988). Readers who are interested are referred to Costa & McCrae (1992) for a recent review of supporting evidence for a biological basis for the “Big Five”. From the practitioner’s point of view all we require 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 Orpheus minor scales

A proper understanding of the Orpheus minor scales demands some familiarity with the literature concerning integrity testing. Integrity tests have been in widespread use in the USA for several years 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 tend to develop their integrity tests in-house, while the 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 are reviewed in the Tenth Mental Measurement Yearbook (Conoley and Kramer, 1989). In addition, the American Psychological Association (Goldberg et al, 1991) and the US Office of Technological Assessment (1990) have both reviewed the use of integrity tests, and these reports are summarised by Camara and Schneider (1994) in the American Psychologist. In the APA survey, less than half of the publishers of integrity tests supplied the information that was requested. Camara and Schneider are concerned that commercial interest has so restricted the dissemination of information concerning the effectiveness of integrity testing that the case for their use is essentially unanswered. They also point out that, among the various instruments available, there is little agreement on the behaviours assessed or on a precise definition of integrity. They find the concept of integrity to be overly broad and ill-defined, and conclude 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) and by Lykken (see Ones et al, 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

TEST SPECIFICATION

(Klump, 1984), the Reid Report (Reid Psychological Systems, 1951), the Phase II profile (Lousig-Nont, 1987), the Milby Profile (Miller and Bradley, 1975) and the Trustworthiness Attitude Survey (Cormack & Strand, 1970). 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, 1972), the Employment Inventory of Personnel Decisions (Paajanen, 1985; PDI, 1985) and the Hogan Personality Inventory (Hogan, 1981). 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 his 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 as a predictor of overall job performance than any of the "Big Five" factors either alone or in combination. In comparison with supervisors ratings of overall job performance they find an unadjusted validity of 0.22 for integrity (0.41 adjusted for attenuation), which compares favourably with the highest validities reported by Tett and Jackson (1991) in their meta-analysis of occupational personality testing where they found a validity of 0.22 for the "Big Five" trait of agreeableness.

O'Bannon et al (1989) lists 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, tardiness and theft among target behaviours for integrity tests. Rust (1996), in the Manual for the Giotto integrity test, summarises the target behaviours within a broad spectrum formulation of integrity based on 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 (1994) 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 not 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

TEST SPECIFICATION

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) and Sackett et al (1995) 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. To disentangle 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 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 integrity by a specific scale designed for this purpose. Interpretation can then be based on a judgement of both of these 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 Orpheus "Big Five" scales. The second approach is used by the seven Orpheus minor 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 his 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' handbook 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, Schmidt and Viswesvaran 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.

TEST SPECIFICATION

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 on the basis of 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 as long-term commitment, laziness and absenteeism, which is the primary argument in favour of using personality tests at all. In the Orpheus seven minor 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, in reality 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 "statistical" 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 his “Purgatorio”, 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 Manual (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 Orpheus minor 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 Orpheus minor scales*

<i>Classical virtues/vices</i>	<i>Integrity trait</i>	<i>Orpheus minor scale</i>
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 Orpheus minor trait of Disclosure maps on to social desirability rather than to the integrity trait of Greed. The Orpheus minor scales assess the reverse of their corresponding integrity trait.

THE CONSTRUCTION OF THE ORPHEUS SCALES

The major scales

The Orpheus scales were constructed progressively and incrementally from core attributes of their traits which had been predefined for items with the use of markers (Golombok & Rust, 1993; 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 in to 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 major 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 minor 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 minor scales, the requirement that each item should be associated with only one minor scale in terms of both reliability and validity kept artificial correlations between the minor seven to a minimum.

Given this development process, how might we expect the five major scales from Orpheus 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

sion 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 impulsively items from his earlier extra-version scale (Claridge & Birchall 1978; Rust, 1975, 1987, 1988, 1989). 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 Orpheus, 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 neer-do-wells that have been implied. Successful senior managers, and human resource managers in particular, 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, Orpheus 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 minor scales were constructed. This has led to an emphasis on the previously obscured merits of low scores on agreeableness and on conscientiousness in the Orpheus Authority and Detail scales respectively.

Within the Orpheus 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 have to 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, in particular 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 also more educated – often at degree level. 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 lower correlations with Authority and by placing the trait in the intellectual domain, Orpheus 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.

Orpheus 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. In spite of 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 Orpheus 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 minor scales

The Giotto integrity test was specifically designed to address statistical problems involved in the assessment of integrity, in particular the non-linearity of many of the fundamental concepts and of social desirability itself. The Orpheus seven minor scales, however, are constructed to the classical psychometric model which assumes linearity. An analysis of cross-correlation between Giotto and Orpheus items showed that the focal point of the resultant discrepancy was the Greed dimension which mapped directly on to the interim Orpheus social desirability scale, itself highly related to social desirability. This interim social desirability scale has therefore been firmed up as one of the seven minor scales described as P6: 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 fairly 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 major scales, the Orpheus seven minor scales explore the space closer to this vortex in mapping onto the Giotto framework.

The Orpheus minor 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

THE CONSTRUCTION OF THE ORPHEUS SCALES

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 P6 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 major 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 similar to 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 Orpheus 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 previously developed by Rust (1976, 1988, 1989), Rust and Golombok (1986) and Golombok and Rust (1994). Each scale had a balanced number of positive and negative items. The five major 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 five major scales had been identified, seven minor scales were assembled from the remaining pilot items. One of these (P6: Disclosure) was assembled from the interim social desirability scale. The final version of Orpheus consisted of 190 items.

THE STANDARDISATION STUDY

Description of the sample

Orpheus 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 males and 138 females, and the mean age of the sample was 30.67 years with a standard deviation of 11.01 years. 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 represents 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 Orpheus. These are used to produce the standardised stanine scores which are used in the narrative report.

Statistical Analysis

The data were standardised within each respondent (Rust & Golombok, 1994). 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 $x = (x - \text{mean}) / \text{s.d.}$ This was carried out in order to eliminate response bias effects resulting from acquiescence. Item analysis, reliability analysis and validation was conducted with data that had been pre-transformed in this way.

Development of Response Audit Scales

Four Response Audit (screening) scales were derived from the standardisation data. These were;

RA1: Dissimulation

RA2: Ambivalence

RA3: Despondency**RA4: Inattention**

Scores on all Response Audit scales range from 0 through to 3, with the majority of respondents scoring 0 on all audits.

RA1: Dissimulation

The RA1: Dissimulation screen was derived from the P6: Disclosure scale (after standardisation but before rounding) as follows:

	Percentage obtaining this RA1 score in the standardisation sample	
P6 < 1.500 s.d	RA1 = 0	88.9
P6 > 1.500 s.d. and < 1.875	RA1 = 1	3.6
P6 > 1.875 s.d. and < 2.250	RA1 = 2	4.6
P6 > 2.250 s.d.	RA1 = 3	2.9

An RA1 score of 3 indicates that the respondent is almost certainly lying deliberately, an RA1 score of 2 indicates a very high probability that the respondent is lying and an RA1 score of 1 suggests that the possibility that the respondent is lying should always be taken into account in interpreting his/her scores.

The scale on which RA1 is based (P6: Disclosure) was itself derived from the pilot social desirability scale which is referred to earlier. 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 take into account other factors in the respondents background when interpreting his or her scores. These will include the circumstances of testing and the background of the client.

RA2: Ambivalence

The RA2: 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

THE STANDARDISATION STUDY

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 with 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 (after exclusion for missing data), 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 RA2, 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 RA2: Ambivalence screen was derived as follows from these raw numbers of contradictory statements:

	Percentage obtaining this RA2 score in the standardisation sample	
5 or less	RA2 = 0	89.3
6	RA2 = 1	8.0
7	RA2 = 2	2.5
8 or more	RA2 = 3	0.2

Any person who obtains a high RA2 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 RA2 score.

It should be noted that the expected score for a “genuine” random responder on RA2 is 1 (6 out of 12 on the binomial distribution), and there is therefore a chance of about 40% that such a random pattern will fail to be detected by the RA2 auditor. 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.

RA3: Despondency

The RA3: Despondency auditor was derived from the P6: Disclosure scale (after standardisation but before rounding) as follows:

	Percentage obtaining this RA3 score in the standardisation sample	
P6 > -1.875 s.d	RA3 = 0	95.7
P6 < -1.875 s.d. and > -1.125 s.d.	RA3 = 1	1.4
P6 < -1.125 s.d. and > -2.500 s.d.	RA3 = 2	1.9
P6 < -2.500 s.d.	RA3 = 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 RA3 are set at more extreme values than those for RA1 as respondents are more likely to wish a positive than a negative bias to their results.

It is very probable that the pattern of Orpheus scores from respondents who obtain a RA3 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 his or her employment environment, adopting a "who cares" attitude. Alternatively, the respondent may be depressed or "crying out for help".

RA4: Inattention

The RA4 auditor 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 STANDARDISATION STUDY

The RA4: Inattention auditor was derived as follows from these raw numbers of changes:

	Percentage scoring at this value in the standardisation sample	
between 89 and 141	RA4 = 0	94.4
between 81 and 88 or between 142 and 149	RA4 = 1	5.2
between 71 and 80 or between 150 and 154	RA4 = 2	0.4
less than 70 or greater than 154	RA4 = 3	0.0

The RA4: 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 major 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 P6: Disclosure scale.

Intercorrelations between the Orpheus major scales are shown in Table 8.

Table 8: Intercorrelations between the Orpheus major scales and correlations with social desirability in the standardisation sample. Social desirability is assessed by the inverse P6 scale, so that that a low score on P6: 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.

	<i>F</i>	<i>A</i>	<i>C</i>	<i>E</i>	<i>D</i>	<i>P6(reversed)</i>
F: Fellowship	(.73)	.19	-.19	-.21	-.03	.05
A: Authority		(.77)	-.28	-.25	-.12	-.05
C: Conformity			(.76)	.15	.28	.11
E: Emotion				(.81)	-.06	-.20
D: Detail					(.73)	.29
P6: Disclosure (reversed)						(.76)

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

Table 9: Numbers of items from each of the major scales used to augment each of the minor scales (*U* = no overlap)

	<i>U</i>	<i>F</i>	<i>A</i>	<i>C</i>	<i>E</i>	<i>D</i>
P1: Proficiency	11	0	0	1	2	0
P2: Work-orientation	15	0	1	1	2	1
P3: Patience	10	0	4	0	0	0
P4: Fair-mindedness	11	1	0	5	2	1
P5: Loyalty	12	0	2	1	1	2
P6: Disclosure	16	0	0	0	0	0
P7: Initiative	12	0	0	1	0	1

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

	P1	P2	P3	P4	P5	P6	P7
P1	(.70)	.26	.44	.19	.35	-.53	.11
P2		(.70)	-.02	.41	-.19	-.27	.47
P3			(.73)	.20	.42	-.29	-.05
P4				(.72)	-.13	-.08	.44
P5					(.73)	-.32	-.45
P6						(.76)	-.12
P7							(.72)

TEST SPECIFICATION

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

	<i>F</i>	<i>A</i>	<i>C</i>	<i>E</i>	<i>D</i>
P1	-.01	-.04	.11\$	-.40\$.55
P2	.15	.49\$	-.38\$	-.31\$.20\$
P3	-.07	-.43\$.03	-.22	.29
P4	.25\$.18	-.62\$	-.32\$.05\$
P5	-.16	-.52\$.41\$.25\$.52\$
P6	-.05	.05	-.11	.20	-.29
P7	.38	.49	-.47\$	-.48	-.07\$

From Table 11 we can see that while both the P1: Proficiency and the P2: Work-orientation scale reflect attitudes to work, high scorers on P1: 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. Meanwhile higher scorers on P2: Work-orientation tend to obtain high scores on Authority and low scores on Conformity, suggesting that they are achievement oriented (see Table 6). High scorers on P1: proficiency also tend to obtain high scores on P3: Patience and P5: Loyalty, reflecting the distracting effects that a lack of patience or of loyalty may have on proficiency. P2: Work-orientation, on the other hand, correlates with P4: Fair-mindedness and P6: Initiative, the latter in particular, suggesting a positive orientation to their working environment. The high negative correlation between P1: Proficiency and P6: Disclosure reflects the fact that a lack of proficiency is normally perceived as particularly socially undesirable. P6 is able to detect a reluctance to disclose socially undesirable attributes. Scores on P6 should always be noted carefully when interpreting P1.

Low scorers on P3: Patience tend to obtain high scores on Authority reflecting the common observation that aggression (low P3) and ambition often go hand in hand. High scorers on P4: 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 P5: 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 P5 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 P7: Initiative reflect the expected 'Big Five' pattern for entrepreneurs (see Table 6), while Table 10 confirms this pattern, showing that while such individuals may be work-oriented and fair-minded, they tend to have low P5: Loyalty scores which indicates a preference for being their own boss. P6: Disclosure has relatively low correlations with all five major scales, confirming the effectiveness of the procedures used to reduce social desirability correlations for the five major scales to below 0.3. The correlations of P6: Disclosure with the P1, P2, P3 and P5 minor 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 Orpheus variables with age, sex and educational level for the 427 respondents in the standardisation sample; (for sex, male = 1, female = 2); for Education a higher score indicates a higher educational level).

	Raw correlation			Pooled within occupations		
	Age	Sex	Education	Age	Sex	Educ.
F:Fellowship	-.13*	.08	-.02	-.21**	.11	-.04
A:Authority	.14**	-.06	.17**	.08	-.04	.13*
C: Conformity	-.08	.05	-.53**	.11	.10	-.34**
E: Emotion	-.02	.14**	-.02	-.01	.10	-.06
D: Detail	.04	.07	-.13*	.03	-.02	-.03
P1: Proficiency	.26**	.09	.02	.13*	.02	.01
P2: Work-orientation	.22**	.09	.20**	.09	.02	.07
P3: Patience	.08	.02	.08	.02	-.06	.02
P4: Fair-mindedness	.27**	.01	.53**	.04	-.17**	.35**
P5: Loyalty	.04	.15**	-.17**	.02	.08	-.11
P6: Disclosure	.10*	-.26**	-.10*	-.19**	-.06	.08
P7: 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 (Golombok and Rust 1993; Rust 1996). 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 Orpheus scale using analysis of variance. No significant differences were found.

RELIABILITY

RELIABILITY

Reliability of the Orpheus scales was calculated using the split-half method from the data in the standardisation study. Details of the reliability for each scale appear in Table 3 of this Manual.

VALIDITY

Content validity of the five major scales of Orpheus

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". As a consequence of this, the item total correlations of these two items with scores on Fellowship will be positive and negative respectively.

F: 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."

A: 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."

C: Conformity

Positive: "Changing the way we do things usually makes matters worse."

Negative: "I am persistently on the lookout for new ideas to exploit."

E: 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."

D: 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."

VALIDITY

Content validity of the seven minor scales of Orpheus

P1: 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."

P2: 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."

P3: 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."

P4: Fair-mindedness

Positive: "People are usually honest with me."

Negative: "Many people are so naive that it's very easy to manipulate them."

P5: Loyalty

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

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

P6: Disclosure

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

Negative: "I always tell the truth."

P7: 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 Orpheus five major 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 minor scales of Orpheus are based on a more recent model and new evidence is therefore required. Some is given in the Giotto Manual which reports a

cross validation of the Prudentius scales with a 94 item adjective check list constructed specifically for the Prudentius model. 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 major 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 carelessness, absenteeism, hostility, subversion, disloyalty, greed and inertia. 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 Orpheus standardisation study. The seven check-list scales are detailed below.

Checklist scale 1 (Carelessness) consisted of 8 adjectives (4 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) consisted of 8 adjectives (4 positive and 4 negative), the highest factor loadings being for Irresponsible and Unkempt (negative) and Determined and Tireless (positive).

Checklist scale 3 (Hostility) 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) 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) 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) 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) 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).

VALIDITY

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

Orpheus scale	Checklist scale						
	1	2	3	4	5	6	7
F: Fellowship	.04	-.09	.15**	-.05	.18**	-.09	-.23**
A: Authority	.02	-.35**	.42**	.02	.47**	.04	-.31**
C: Conformity	-.11*	.27**	-.15*	.34**	-.28**	.22**	.43**
E: Emotion	.04	.23**	-.01	.14**	-.12*	.08	.46**
D: Detail	-.43**	-.15**	-.24**	.02	-.38**	-.05	.13*
P1: Proficiency	<u>-.39**</u>	-.24**	-.26**	-.17**	.29**	-.16**	-.11*
P2: Work-orientation	-.14**	<u>-.48**</u>	.18**	-.19**	.13**	-.14**	-.41**
P3: Patience	-.12*	.06	<u>-.54**</u>	-.21**	.39**	-.18**	-.03
P4: Fair-mindedness	-.05	-.32**	-.08	<u>-.48**</u>	.05	-.38**	-.39**
P5: Loyalty	-.25**	.12*	-.43**	-.05	<u>-.56**</u>	-.06	.28**
P6: Disclosure	.26**	.17**	.17**	.11*	.23**	<u>.20**</u>	.11*
P7: Initiative	.05	-.32**	.22**	-.20**	.36**	-.17**	<u>-.55**</u>

The correlations between the seven checklist scales and the seven Orpheus minor 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 Orpheus 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 P6: Disclosure, a robust measure of socially desirable response. This is shown in Table 14.

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

Orpheus scale	Checklist scale						
	1	2	3	4	5	6	7
F: Fellowship	.06	-.08	.17**	-.05	.20**	-.08	-.23**
A: Authority	.00	.37**	.42**	.01	.47**	.03	-.33**
C: Conformity	-.07	.32**	-.12*	.37**	-.25**	.27**	.46**
E: Emotion	-.03	.20**	-.06	.11*	-.18**	.03	.45**
D: Detail	-.37**	-.10*	-.19**	.07	-.33**	.03	.18**
P1: Proficiency	-.37**	-.18**	-.21**	-.13*	.20**	-.05	-.06
P2: Work-orientation	-.08	-.46**	.24**	-.16**	.20**	-.10	-.39**
P3: Patience	-.04	.12*	-.52**	-.19**	-.34**	-.12*	.00
P4: Fair-mindedness	-.03	-.31**	-.07	-.47**	.07	-.38**	-.38**
P5: Loyalty	-.17**	.20**	-.40**	-.02	-.52**	.01	.34**
P7: Initiative	.08	-.31**	.24**	-.19**	.40**	-.15**	-.54**

It can be seen in Table 14 that partialling out of the social desirability effect eliminates some of the smaller significant correlations, but leaves most unchanged. The pattern of intercorrelations found for the 'Big Five' conform 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 minor 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 minor 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 partialling out the effect of all five major scales from the correlation matrix between Orpheus and the Prudentius checklist scales. This partial correlation matrix is shown in Table 15.

VALIDITY

Table 15: The partial correlations of the seven Prudentius checklist scales with the seven Orpheus minor scales, with the social desirability scale (P6: Disclosure) and the five major scales as partial variates

Orpheus scale	Checklist scale						
	1	2	3	4	5	6	7
P1:Proficiency	<u>-.16**</u>	-.13*	-.13*	-.17"	-.10	-.07	-.03
P2: Work-orientation	-.01	<u>-.22**</u>	.10	-.07	-.02	-.03	-.19**
P3: Patience	.02	.06	<u>-.40**</u>	-.16**	-.15**	-.09	-.03
P4: Fair-mindedness	-.03	-.10	-.19**	<u>-.33**</u>	-.12*	-.28**	-.04
P5:Loyalty	-.01	.01	-.20**	-.18	<u>-.23**</u>	-.05	-.03
P7: Initiative	.06	-.04	.05	-.05	.15**	-.08	<u>-.23**</u>

On the basis of this analysis the following evidence for the validity of the minor scales is suggested:

P1: Proficiency. The correlation with adjectival Carelessness remains significant. This suggests that the P1 scale is valid as a predictor of Carelessness, and provides data on this trait over and above that already implicit in the Detail scale. Other correlations of P1 with Checklist scales 2, 3 and 4 are also supportive of the validity of P1.

P2: Work-orientation. The negative correlation with Checklist scale 2 supports the validity of P2 as a Work-orientation scale. The correlation of P2 with checklist scale 7 provides some additional support for validity of P2.

P3: Patience. The negative and very significant correlation of Patience with Checklist scale 3 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.

P4: Fair-mindedness. The correlation of P4 with Checklist scale 4 is extremely encouraging. The correlation with Checklist scale 6 is also very supportive of the validity of P4.

P5: Loyalty. Evidence for the validity of this scale is provided by the significant correlation with Checklist scale 5.

P7: Initiative. The significant correlation with Checklist 7 is consistent with many of the negative attributes of P7. P7 is the minor scale that relates most strongly to the major scales, albeit one which involves a complex interpretation of the profile from four of the five. It is encouraging to see that P7 is indeed adding something over and above the combination of High Fellowship, High Authority, Low Conformity and Low Emotion which on their own would represent an Entrepreneurship profile.

Criterion-related validity

The five Orpheus major scales were validated against 10 predesignated supervisors rating scales (one positive and one negative rating for each scale). Ratings ranged from below average through to 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.

VALIDITY

Table 16: *Correlations of the five Orpheus major scales with supervisors ratings (N=200 (approximately))*

	F: Fellowship
Team skills	.19**
Ability to work independently	-.14*
	A: Authority
Ability to make friends with colleagues	-.13+
Ability to make tough decisions	.24**
	C: Conformity
Ability to generate new ideas	-.25**
Obedience to company policy	.23**
	E: Emotion
Level of self-confidence	-.25**
Tendency to worry	.08
	D: Detail
Attention to detail	.23**
Breadth of vision	-.22**

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

+ = $p < .05$ (one tailed)

Note that the validity coefficients that appear in Table 16 are not adjusted for attenuation and therefore represent high values for validity coefficients of this type.

Because of the nature of the minor scales it is not appropriate to pool the supervisors ratings for all the respondents in the validation study as what counts as appropriate behaviour in so far as these scales is concerned varies considerably from occupation to occupation. However, a correlational analysis relating supervisor's ratings to scores on the minor 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 Orpheus minor scales.

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

	P1: Proficiency
Time-keeping	.33*+
Performance under pressure	.27*
	P2: Work-orientation
Attention to detail	.28*
	P3: Patience
Level of self-control	.26*
Dependability	.27*
	P4: Fair-mindedness
Trustworthiness	.29*
Ability to work to strict guidelines	.26*
	P5: Loyalty
Obedience to company policy	.26*
Ability to make friends with colleagues	.26*
	P6: Disclosure
Likelihood of making mistakes	-.27*
	P7: Initiative
Level of self-confidence	.28*

* = $p < .05$, *+ = $p < .02$

—

—

—

—

REFERENCES

- Adler A (1964) *Social interests: A challenge to mankind*. Capricorn Books; New York.
- Adorno T W, Frenkel-Brunswick, Levinson D J & Sanford R N (1950) *The Authoritarian Personality*. Harper & Row; New York.
- Allport G W & Odbert H S (1936) Trait-names: A psycholexical study. *Psychological Monographs*, **47**, (1, Whole No. 211).
- Almagor M, Tellegen A & Wailer 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 A (1990) *Psychological testing* (6th edition). Macmillan; New York.
- Angleinter A, Ostendorf F & John O P (1990) Towards a taxonomy of personality descriptors in German: A psycho-lexical study. *European Journal of Personality*, **4**, 89–118.
- Atkinson J W (1977) Motivation for achievement. In T Blass (Ed) *Personality variables in social behaviour*. Lawrence Earlbaum; New York.
- Atkinson J W & Birch D (1970) *The dynamics of action*. Wiley; New York.
- Atkinson J W & Birch D (1974) The dynamics of achievement oriented activity. In W Atkinson & J Raynor (Eds) *Motivation and achievement*. Winston. Washington DC.
- Avia D, Sanz J, Martinez-Arias R, Silva F & Grana J (1995) The five factor model-II: Relations of the NEO-PI with other personality variables. *Personality and Individual Differences*, **19**, 81–97.
- 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.
- Briggs R S (1992) Assessing the five factor model of personality description, *Journal of Personality*, **60**, 253–293.
- Buss D & Cantor N (1989) *Personality psychology: Recent trends and emerging directions*. Springer-Verlag; New York.

REFERENCES

- Camara W J & Schneider D L (1994) Integrity tests: Facts and unresolved issues. *American Psychologist*, **49**, 112-119
- Caprara V V & Perugini M (1994) Personality described by adjectives: Generalizability of the "Big Five" to the Italian lexical context. *European Journal of Personality*, **8**, 357-369.
- 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 Eber H W & Tatsuoka M M (1970) *Handbook for the sixteen personality factor questionnaire* (3rd edition) Institute for Personality and Ability Testing; Champaign, Illinois.
- Cattell R B (1995) The fallacy of the five factors in personality. *The Psychologist*, May, 207-208.
- Church T A & Katigbak M S (1989) Internal, external and self report structure of personality in a non-western culture: An investigation of cross language and cross-cultural generalizability. *Journal of Personality and Social Psychology*, **57**, 857-872.
- 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.
- Cormack R W & Strand A L (1970) *Trustworthiness Attitude Survey*. Personnel Systems Corporation; Oakbrook IL.
- Costa Jr P T (1991) Clinical use of the five factor model. *Journal of Personality Assessment*, **57**, 393-396.
- Costa Jr P T, Bush C M, Zonderman A B & McCrae R R (1986) Correlations of MMPI factor scales with measures of the five-factor model of personality. *Journal of Personality Assessment*, **50**, 640-650.

Costa P T & McCrae R R (1995) Primary traits of Eysenck's P-E-N system: Three- and five- factor solutions. *Journal of Personality & Social Psychology*, **69**, No. 2, 308-317.

Costa Jr P T & McCrae R R (1988) Personality in adulthood: A six year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, **54**, 853-863.

Costa Jr P T, McCrae R R & Dembroski T M (1989) Agreeableness vs. Antagonism: Explication of a potential risk factor for CHD. In A Siegman & T M Dembroski (Eds) *In search of coronary-prone behaviour: Beyond Type A*. Lawrence Earlbaum Associates; Hillsdale NJ.

Costa Jr P T, Stone S, McCrae R R, Dembroski T M & Williams R B (1987) Hostility, agreeableness-antagonism and coronary heart disease. *Holistic Medicine*, **2**, 161-167.

Costa P T & McCrae R R (1992) Four ways five factors are basic. *Personality and Individual Differences*, **13**, 653-665.

Couch A & Keniston K (1960) Yeasayers and naysayers: Agreeing response set as a personality variable, *Journal of Abnormal and Social Psychology*, **60**, 151-174.

Crowne D P & Marlowe B R (1960) A new scale of social desirability independent of psychopathology, *Journal of Consulting Psychology*, **24**, 349-354.

Dembroski T M, McDougall J M, Costa Jr P T & Grandits G (1988) Components of hostility as predictors of sudden death and myocardial infarction in the multiple risk intervention trial. *Psychosomatic Medicine*, **51**, 514-522.

De Raad B & Szirmak Z (1994) The search for the "Big Five" in a non-Indo-European language: The Hungarian trait structure and its relationship to EPQ and PTS. *European Review of Applied Psychology*, **44**, 17-24.

Digman J M (1989) Five robust trait dimensions: Development, stability, utility, *Journal of Personality*, **57**, 195-214.

Digman J M (1990) Personality structure: Emergence of the five factor model. *Annual Review of Psychology*, **41**, 417-440.

Digman J M & Inouye J (1986) Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology*, **50**, 116- 123.

REFERENCES

- Digman J M & Takemoto-Chock N K (1981) Factors in the natural language of personality: Re-analyses, comparison and interpretation of six major studies. *Multivariate Behavioral Research*, **16**, 149–170.
- Edwards A L (1957) *The social desirability variable in personality assessment and research*. Dryden Press; New York.
- Erikson E H (1950) *Childhood and Society*. Norton; New York.
- Eysenck H J (1947) *Dimensions of personality*. Routledge and Kegan Paul; London.
- Eysenck H J (1970) *The structure of human personality* (3rd edition). Methuen; London.
- Eysenck H J (1992) Four ways in which five factors are not basic. *Personality and Individual Differences*, **6**, 667–673.
- Eysenck H J & Eysenck M W (1975) *Manual of the Eysenck Personality Questionnaire*. San Diego; Educational and Industrial Testing Service.
- Fiske D W (1949) Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, **44**, 329–344.
- Freud S (1957) On narcissism: An introduction. In J Strachey (Ed) *The standard edition of the complete psychological works of Sigmund Freud*. Hogarth Press; London.
- Furnham, A (1986) Response bias, social desirability and dissimulation. *Personality and Individual Differences*, **7**, 385–406.
- Furnham A (1990) Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology: Research and Reviews*, **9**, 46–55.
- Furnham A & Henderson M (1982) The good, the bad and the mad: response bias in self report measures. *Personality and Individual Differences*, **3**, 311–320.
- Galton F (1884) Measurement of character. *Fortnightly Review*, **42**.
- Goldberg L R (1981) Language and individual differences: the search for universals in personality lexicons. In L Wheeler (Ed) *Review of Personality and Social Psychology*. Sage; Beverley Hills.

Goldberg L R (1982) From Ace to Zombie: Some explorations in the language of personality. In C D Spielberger & J N Butcher (Eds) *Advances in Personality Assessment*. Lawrence Earlbaum; Hillsdale NJ.

Goldberg L R (1990) An alternative "Description of Personality". The big five structure. *Journal of Personality and Social Psychology*, **59**, 1216-1229.

Goldberg L R (1992) The development of markers of the big five structure. *Psychological Assessment*, **4**, 26-42.

Goldberg L R (1993) The structure of phenotypic personality traits. *American Psychologist*, **48**, 26-34.

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

Golombok S & Rust J (1992) *The Manual for the Third Revision of the Wechsler Intelligence Scale for Children, UK Edition (WISC-III (UK))*. The Psychological Corporation; London.

Golombok S & Rust (1993) The measurement of gender role behaviour in pre-school children. *Journal of Child Psychology and Psychiatry*, **34**, 805-811.

Golombok S & Rust J (1993) *The Pre-School Activities Inventory*. *Psychological Assessment*, **5**, 131 - 136.

Gough H G (1971) The assessment of wayward impulse by means of the Personal Reaction Blank. *Personnel Psychology*, **24**, 669-677.

Gough H G (1972) *Personnel Reaction Blank*. Consulting Psychologists Press; Palo Alto CA.

Gough H G (1987) *California Psychological Inventory administrator's guide*. Consulting Psychologists Press; Palo Alto CA.

Graziano W G & Eisenberg N H (1993) Agreeableness: A dimension of personality. In S R Briggs, R Hogan & W H Jones (Eds.) *Handbook of Personality Psychology*. Academic Press; New York.

Guilford J P & Zimmerman W S (1949) *The Guilford-Zimmerman Temperament Survey*. Sheridan Supply; Beverley Hills CA.

REFERENCES

- 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.
- Hogan J & Hogan R (1989) How to measure employee reliability. *Journal of Applied Psychology*, **74**, 273-279.
- Hogan R T (1991) Personality and personality measurement. In M D Dunnette & L M Hough (Eds) *Handbook of industrial and organizational psychology (Vol 2)*. Palo Alto, CA; Consulting Psychologists Press.
- Hartshorn H, May M A & Maller J B (1929) *Studies in the nature of character*. Macmillan; New York.
- Homey K (1945) *Our inner conflicts*. W W Norton; New York.
- Holland J L (1985) *Manual for the Vocational Preference Inventory*. Odessa, FL; Psychological Assessment Resources.
- Holland J L (1985) *Making vocational choices: A theory of vocational personalities and work environments*. Englewood Cliffs, NJ; Prentice-Hall.
- James W (1907) *Pragmatism: A new name for some old ways of thinking*. Longmans Green; London.
- John O P (1989) Towards a taxonomy of personality descriptors. In D Buss & N Cantor (Eds) *Personality psychology: Recent trends and emerging directions*. Springer-Verlag; New York.
- 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.
- Kashiwagi S, Wada S & Aoki T (1993) The big five and the oblique primary pattern for the items of the ACL, Japanese version. *Japanese Journal of Psychology*, **64**, 153-159.
- Klump C S (1964) *Stanton Survey*. Stanton Corporation; Charlotte NC.
- Knowles J B (1963) Acquiescence response set and the questionnaire measurement of personality. *British Journal of Social and Clinical Psychology*, **2**, 131 - 137.

- Kohnstamm A G, Mervielde I, Besevegis E & Halverson Jr F C (1995) Tracing the big five in parents' free descriptions of their children. *European Journal of Personality*, **9**, 283-304.
- 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.
- Lorr M (1986) *Manual of the Interpersonal Styles Inventory*. Western Psychological Services; Los Angeles.
- Lousig-Nont G M (1987) *Phase II Profile*. Lousig-Nont and Associates; Las Vegas, NV.
- McCrae R R (1987) Creativity, divergent thinking and openness-to-experience, *Journal of Personality and Social Psychology*, **52**, 1258-1265.
- McCrae R R (1989) Why I advocate the five factor model: joint factor analysis of the NEO-PI with other instruments. In D M Buss & N Cantor (Eds) *Personality psychology: Recent trends and emerging directions*. Springer-Verlag; New York.
- McCrae, R R & Costa, P T (1985) Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, **49**, 710-721.
- McCrae, R R & Costa, P T (1986) Clinical assessment can benefit from recent advances in personality psychology. *American Psychologist*, **41**, 1001-1003.
- McCrae R R & Costa Jr P T (1987) Validation of the five factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, **52**, 81-90.
- McCrae R R & Costa Jr P T (1989) Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality, *Journal of Personality*, **57**, 17-40.

REFERENCES

- McCrae R R & Costa Jr P T (1993) Conceptions and correlates of openness-to-experience. In S R Briggs, W H Jones & R Hogan (Eds) *Handbook of Personality*. Academic Press; New York.
- McCrae R R & Costa Jr P T (1995) Trait explanations in personality psychology. *European Journal of Personality*, **9**, 231–252.
- McCrae R R, Costa Jr P T & Piedmont R L (1992) Folk concepts, natural language and psychological constructs: The California Personality Inventory and the five factor model. *Journal of Personality*, **61**, 1–26.
- McCrae, R R & John, O P (1992) An introduction to the Five-Factor model and its applications, *Journal of Personality*, **60**, 175–215.
- McClelland D C, Atkinson J W, Clark R A & Lowell E L (1953) *The achievement motive*. Appleton-Century-Crofts; New York.
- 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.
- Miller T R (1991) The psychotherapeutic utility of the five factor model of personality: A clinician's experience, *Journal of Personality Assessment*, **57**, 415–433.
- Murray H A (1938) *Explorations in personality*. Cambridge; Harvard University Press.
- Myers I B & McCauley M H (1985) *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Consulting Psychologists Press; Palo Alto CA.
- Norman W T (1963) Towards an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, **66**, 574–583.
- Norman W T (1967) *2800 personality trait descriptors: Normative operating characteristics for a university population*. Ann Arbor, University of Michigan, Department of Psychology.
- Norman W T & Goldberg L R (1966) Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, **4**, 681–691.

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.

Ostendorf F & Angleiter A (1994) A comparison of different instruments proposed to measure the big five. *European Review of Applied Psychology*, **44**, 45–53.

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

Passini F T & Norman W T (1966) A universal conception of personality structure? *Journal of Personality and Social Psychology*, **4**, 44–49.

Paulhus D L (1984) Two-component model of socially desirable responding. *Journal of Personality and Social Psychology*, **46**, 598–609.

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; New York.

Personnel Decisions Incorporated (1985) *Development and validation of the PDI Employment Inventory*. Author Minneapolis, MN.

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.

REFERENCES

- Pulver A, Allik J, Puikkinen L & Hamalainen M (1995) A big five personality inventory in two non-European languages. *European Journal of Personality*, **9**, 109–124.
- Reid Psychological Systems (1951) *Reid Report*. Author Chicago.
- Rokeach M (1954) The nature and meaning of dogmatism. *Psychological Review*, **61**, 194–204.
- Rokeach M (1960) *The open and closed mind*. Basic Books; New York.
- Rotter J B (1966) Generalized expectancies for internal versus external locus of control or reinforcement. *Psychological Monographs*, **80**(1).
- Rust J (1974) Interaction of reliabilities in personality measurement. *Social Behaviour and Personality*, **2**, 108–110.
- Rust J (1975) Genetic effects in the cortical auditory 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 (1976) Generalisation and dishabituation of the orienting response by a stimulus of lower intensity. *Physiological Psychology*, **4**, 99–101.
- Rust J (1976) Unconditioned response diminution in the skin resistance response. *Journal of General Psychology*. **95**, 77–84.
- Rust J (1977) Habituation of the orienting response in the auditory cortical evoked potential. *Psychophysiology*, **14**, 123–126.
- 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) Sociobiology and psychometrics: Do they really need each other? *Philosophical Psychology*, **1**, 117–129.
- 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 & Golombok S (1986) *Manual of the Golombok Rust Inventory of Sexual Satisfaction (GRISS)*. NFER-Nelson; Windsor.

Rust J, Bennun I, Crowe M & Golombok S (1988) *Manual of the Golombok Rust Inventory of Marital State (GRIMS)*. NFER-Nelson; Windsor.

Rust J, Moncada A & LePage B (1988) Personality dimensions through the schizophrenia borderline. *British Journal of Medical Psychology*, **61**, 163–166.

Rust J & Chiu H. (1988) Schizotypal estimators in adolescence. *Social Behaviour and Personality*, **16**, 25–31.

Rust J, Golombok S and Abram M (1989) Creativity and schizotypal thinking. *Journal of Genetic Psychology*, **150**, 225–227.

Rust J & Golombok S (1994) *Modern Psychometrics*. Routledge; London.

Rust J, Golombok S & Trickey G (1993) *Manual of the Wechsler Objective Reading Dimensions (WORD)*. The Psychological Corporation; London.

Rust J (1996) *Manual of the Wechsler Objective Numerical Dimensions (WOND)*. The Psychological Corporation; London.

Rust J (1996) *Manual of the Wechsler Objective Language Dimensions (WOLD)*. The Psychological Corporation; London.

Rust J (1996) *Manual for the Wechsler Quick Test*. The Psychological Corporation; London.

Rust J (1996) *Manual for the Wechsler Individual Achievement Test (UK Edition)*. The Psychological Corporation; London.

Rust J (1996) *Giotto Handbook*. The Psychological Corporation; London.

Sackett P R, Burns L R & Callahan C (1989) Integrity testing for personnel selection: An update. *Personnel Psychology*, **37**, 491–529.

Saucier G (1992) Benchmarks: Integrating affective and interpersonal circles with the big five personality factors. *Journal of Personality and Social Psychology*, **62**, 1025–1035.

1

2

3

4