PERGAMON

Personality and Individual Differences 27 (1999) 755–768

PERSONALITY AND INDIVIDUAL DIFFERENCES

The validity of the Giotto integrity test

John Rust*

Goldsmiths College, University of London, New Cross, London SE14 6NW, U.K.

Received 19 July 1998; received in revised form 2 December 1998; accepted 8 December 1998

Abstract

Giotto is a work-based personality questionnaire designed to assess integrity. It contains 101 items and generates scores on seven scales based on the Prudentius model of personality. Because these traits are evaluative in nature, Giotto makes use of an ipsative format in which force choice items are matched for social desirability. The intrinsic non-linearity of the ipsative framework is addressed using neural network programming techniques. Giotto has been standardised on 701 people in a variety of employment positions. Several studies supporting the validity of Giotto are reported. © 1999 Elsevier Science Ltd. All rights reserved.

1. Introduction

Sackett, Burris, and Callahan (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'). Disguised purpose integrity tests differ from other personality tests only in terms of the nature of the traits being assessed. Conoley and Kramer (1989), the US Congress, Office of Technology Assessment (1990), Goldberg, Grenier, Guion, Sechrest, and Wing (1991), Camara and Schneider (1994) and Sackett and Wanek (1996) have reviewed the use of these tests.

The case in favour of integrity testing is made by Deniz Ones and her colleagues (Schmidt, Ones, & Hunter, 1992; Collins & Schmidt, 1993; Ones, Viswesvaran, & Schmidt, 1993; Ones, Viswesvaran, & Schmidt, 1995). They report a series of meta-analytic studies that review the evidence for the validity of integrity testing on the basis of 650 criteria-related validity coefficients from over 500,000 subjects. They conclude that the evidence for the validity of integrity tests is substantial and that the broad construct of integrity is probably as good or better a predictor of overall job performance than any one of the Big Five personality factors (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991; Schmidt & Ryan, 1993) either alone or in combination. In comparison with supervisors' ratings of overall job performance they find an unadjusted validity of 0.22 for integrity

* Tel. +44-171-919-7886; fax: +44-171-919-7886; e-mail: j.rust@gold.ac.uk

0191-8869/99/$ - see front matter
PII: S0191-8869(98)00277-3

(0.41 adjusted for attenuation). This compares favourably with the highest validities reported by Tett et al. 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, Goldinger, and Appleby (1989) list some of the constructs used in integrity tests. These include responsibility, long-term job commitment, consistency, proneness to violence, moral reasoning, hostility, work ethics, dependability, depression and energy level. Ones et al. (1993) include disciplinary problems, violence on the job, excessive absenteeism and tardiness among target behaviours for integrity tests. Other integrity characteristics summarised by Schmidt et al. (1992) include drug abuse and theft, as well as the more traditional personality traits of reliability, conscientiousness, adjustment, trustworthiness and sociability. Sackett and Wanek (1996) also mention wasteful use of company time, failure to report theft by others and waste and damage to company materials as possible candidates. They also point out the very broad range of severity of the traits covered by integrity tests that can include not only the rather infrequent criminal activities of theft of money and major fraud but also activities such as 'time theft'.

How do these traits of integrity differ from those of more traditional psychometric personality tests? Superficially, the integrity traits appear rather more 'pre-scientific' than modern conceptions such as conscientiousness or neuroticism. However, this may be more apparent than real. Many of the ideas of modern personality theory, such as Eysenck's extraversion and neuroticism (Eysenck, 1947), owe their origins to ideas first explored in the classical period of Greece and Rome over two thousand years ago. Even theories based on relatively recent techniques such as factor analysis are dependent on the inter-relationships between the 'natural language trait descriptors' of human characteristics listed by Allport and Odbert (1936). They listed about 18,000 words in four categories, personal traits, temporary moods or activities, judgements of personal conduct and capacities and talents. However, before Galton (1884) brought these natural language personality descriptors to psychology and prior to the advent of mathematical statistics, natural language personality descriptors had always been a subject of interest to classical scholars and were considered in many texts from the classical period. One of the most influential of these was the Psychomachia of Prudentius (348 AD–405 AD) (Bergman, 1926; Thompson, 1949). This model was later adapted by theologians and thinkers throughout Christianity and Islam, and the traits concerned became knows as the vices (passions) and virtues (sentiments). While there are far more lexical personality descriptors in English today than there were in Latin of the 4th century AD, the conceptual framework established by Prudentius underpins many of the relationships between them.

Aurelius Prudentius Clementis lived between 348 and 405 AD in the Roman city of Caesaraugusta, later to become the Spanish city of Saragossa. He was a scholar and a man of letters in the employment of the Consul for the Salia region of the Roman Empire. Prudentius authored several manuscripts, including the Apotheosis, the Hamartigenia and the Psychomachia. However, it is the Psychomachia, meaning literally 'battle for the mind', that has been the most influential of his works. It describes the development of personality throughout the four stages of life in terms of a continuous battle between human rationality and basic animal nature, the various aspects of each being personified by warriors in a battle. In his words "In our hearts conflicting affections fight hard in successive combats, and savage war rages hotly within our bones. Man's complex nature is an uproar or rebellion". However, in spite of this conflict, there is an underlying trend towards a resolution in which reason will ultimately reign supreme. Prudentius divided his battling

Table 1
The test specification for Giotto is provided by a mapping of the classical theory of the Psychomachia of Prudentius onto the major integrity traits

Classical virtues/vices	Integrity trait
Prudence/folly	competence versus carelessness (Rust, 1996)
	responsibility, consistency (O'Bannon et al., 1989)
	reliability, conscientiousness (Schmidt et al., 1992)
Fortitude/inconstancy	work-orientation versus absenteeism (Rust, 1996)
	long-term job commitment, work ethic (O'Bannon et al., 1989)
	excessive absenteeism and tardiness (Schmidt et al., 1992)
Temperance/anger	patience versus hostility (Rust, 1996)
	proneness to violence, hostility (O'Bannon et al., 1989)
	violence on the job (Schmidt et al., 1992)
Justice/injustice	fair-mindedness versus subversion (Rust, 1996)
	moral reasoning (O'Bannon et al., 1989)
	disciplinary problems (Schmidt et al., 1992)
Faith/infidelity	loyalty versus disloyalty (Rust, 1996)
	dependability (O'Bannon et al., 1989)
Charity/envy	openness versus disclosure (Rust, 1996)
	trustworthiness (O'Bannon et al., 1989)
	theft, drug abuse (Schmidt et al., 1992))
Hope/despair	initiative versus inertia (Rust, 1996)
	energy level and depression (O'Bannon et al., 1989)

entities into two armies, which he referred to as the vices (Vitium) and the virtues (Virtute), although sometimes he uses the terms Sensus (the sentiments) and Furores (the passions). The Psychomachia was the inspiration for the Divine Comedy of Dante Alighieri (1263–1321) (Sayers, 1955).

Giotto di Bondoni (1267–1377) was one of the many artists and poets of his period who was influenced by the work of Prudentius. In his frescoes at the Arena Chapel in Padua, completed in 1305 AD, he depicts the entities as justice/injustice, hope/despair, charity/envy, faith/idolatry, temperance/anger, fortitude/inconstancy and prudence/folly. The test specification for Giotto makes use of this structure as a framework for the traits specified in the integrity testing literature. This is shown in Table 1.

2. Method

2.1. Test development

A thesaurus was utilised to generate a list of 234 adjectives based on the synonyms and antonyms of the concepts provided by the Prudentius model. The adjectives were each evaluated in terms of their suitability and relevance to work settings. These adjectives were then subjected to a review process in order to address issues of content validity. Teams of human resource managers and

occupational psychologists were asked to review the items in the light of behaviours, attitudes and beliefs perceived to be appropriate for a variety of occupational settings. Recommendations for refinement or deletion were made according to the procedures for content validation recommended by Haynes, Richard, and Kubany (1995). Following review, there were 168 items.

A two-part questionnaire was constructed from the adjectives. Part A consisted of paired items in which respondents were asked to choose which of the two most closely applied to them. This ipsative format was utilised in order to conteract bias that would result from the intrinsic social desirability of virtues (contra vices). Thus, each virtuous item was paired with another virtuous item, so that the respondent was asked to choose between two virtues. Conversely with the items that represented 'vices'. Part B of the questionnaire consisted of single items each with a multiple choice response format (Rust, 1989; Rust & Golombok, 1999).

For each of the seven target traits provided by the model there were equal numbers of positive and negative adjectives, representing virtues and their opposite vices. To achieve the ipsative comparison, each item was paired with one other item of the same polarity (positive or negative) but from a different target trait. There were 24 adjectives (12 positive and 12 negative) from each trait. These 24 items were additionally counterbalanced such that there were 4 items (2 positive and 2 negative) representing a pairing with each of the other 6 traits. If we designate the traits as a, b, c, d, e, f and g, then the following pairings 21 pairings are possible: ab, ac, ad, ae, af, ag, bc, bd, be, bf, bg, cd, ce, cf, cg, de, df, dg, ef, eg and fg. There were 4 such pairings for each combination, yielding 84 item pairs. These were then piloted on a group of 43 subjects, who were asked to indicate which of the adjectives in each pair most applied to them. Responses were item analysed to obtain extremity scores for each item pair. Extremity scores were calculated as follows.

Extremity score = proportion of subjects choosing first choice/Number of subjects

It was assumed that the mean of the extremity scores for each item pair would indicate the relative degree of social desirability associated with each item in the pair. Adjectives were then reassigned to pairs on the basis of their extremity scores. This procedure was designed to provide a set of items for the pilot version in which the adjectives were approximately balanced for social desirability.

2.2. *The pilot version*

Each pair of positive items was then placed with a pair of negative items, subject to the proviso that the four combined items (2 positive and 2 negative) should all be from different target traits. Each of these four item sets was designated an item within the pilot version, making 42 part A items in all. The balanced use of both positive and negative adjectives within each item was designed to alleviate a mood effect whereby respondents could develop an increasing negative mood during testing as a result of being presented with a series of negative adjectives.

Within each 4 adjective item, the respondent would be asked to cross out the adjective which least applied to them (representing an indirect endorsement of the remaining negative item). They would also be asked to assert agreement to the adjective which most applied to them (representing an indirect disclaimer to the other positive item). It was assumed that the deleted adjective would generally be from the negative pair, while the ticked adjective would be from the positive pair

(This was subsequently confirmed). In the event that a positive adjective was deleted or a negative adjective affirmed, this would be scored as a further endorsement or disclaimer for this adjective.

For Part B of the pilot version, 84 of the adjectives that had lower extremity scores were placed randomly in a questionnaire (adjective checklist) format. Twelve items were chosen for each target trait (6 positive and 6 negative) providing 84 items (the remaining 10 being pre-selected marker items).

2.3. Standardisation

The pilot version of Giotto was administered to 701 respondents in a variety of occupations within over 20 companies in the UK. Included were insurance and accountancy staff (including claims negotiators and underwriters), teachers, trainers, security guards, fire fighters, cashiers, drivers, managers, clerical staff and human resource personnel. The sample included 480 males and 206 females, with mean age 36.15 years (S.D. = 13.00 years). The educational level of the respondents was broadly based ranging from zero qualifications to Ph.D. Thirty-six of the respondents did not classify their ethnicity as 'white', this representing the approximate proportion of ethnic minority groups within the UK working population. Five hundred and four of the respondents in the Giotto pilot study were also administered another personality questionnaire (Orpheus, Rust, 1996) that gives scores on the Big Five personality traits as well as on the Prudentius scales and a social desirability scale.

2.4. Item analysis

Target criteria for scale construction by item analysis were (i) reliability of greater than 0.7, (ii) intercorrelations of less than 0.5 and (iii) correlations of less than 0.3 with the Orpheus lie scale. In the first stage of scale development a factor analysis was carried out on the Part B adjectives and seven interim B scales were constructed from the items in each content area of the test specification. Equal numbers of positive and negative adjectives were used for each scale. These interim B scales were used as templates for interim A scales.

The ipsativity of the Part A adjectives is conceptually equivalent to a ranking in order of the seven traits. The correlation between ranks with this design is -0.167. Thus the Part A data set inherits a design-dependent intercorrelation of -0.17. (This compares with one of zero where there is no ipsativity). That is, the expected inter-correlation between variables for this data set where there is no relationship between the variables is -0.17. As factor analysis assumes independence between elements its use is not appropriate for ipsative data of this type.

The complete data set from Part A was used as input and a neural network was trained to distinguish the seven traits established by the interim scales from Part B. The network used was a multi-layer perceptron using a back propagation supervised training algorithm (DTI, 1994) using the Neuralworks Predict Program (Neuralware, 1995). It was shown that the data in part A could predict all seven Part B interim scales. When additional hidden layers were added to the network no significant additional gain in predictive power was obtained. This procedure is algebraically equivalent to a test of goodness-of-fit for a linear model. Thus, there is no evidence that non-linearity resulting from ipsativity has confounded classical test construction statistics when applied to these data. The correlation between predicted and actual interim Part B scales, using validation

data (data from 140 respondents held back from the train and test process used by the neural network), was in excess of 0.5 in all cases. It was concluded that the ipsative data set was consistent with the model.

Items for the interim A scales were selected separately from each domain using the same neural network procedure. For each trait, a network was trained to predict the interim B scale for its domain, using all of the items from part A that included an adjective from that domain. The model was tested using a separate validation data set. Actual/predicted scale correlations ranged between 0.45 and 0.69. Items were selected for the appropriate interim A scale if their input weights exceeded a local criterion in the anticipated direction. The interim A scales were then constructed using simple addition (or subtraction for negative items) of the chosen Part A items and the psychometric properties of both sets of interim scales were examined. As expected, the interim B scales had high internal consistency, but tended to be more highly correlated with the lie (social desirability) scale, while the interim A scales had lower internal consistency but were largely independent of the lie scale.

For each of the seven domains the two interim scales were then combined and factor analysis was used for further item selection. During this selection preference was given to items from Part A, so that only Part B items were available for deletion. The item pool for each domain was reduced until the internal consistency on the validation sample was approximately 0.72. Each of the seven emerging scales was further refined to two additional criteria: (a) no inter-scale correlation should be greater than 0.5 and (b) no correlation with the lie scale should be greater than 0.3.

3. Results

3.1. Reliability

The intercorrelations and reliabilities of the resultant Giotto scales are given in Table 2.

3.2. Correlations with sex, age and educational level

Correlations were carried out between the Giotto scales and sex, age and educational level where this information was available. 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 3.

Looking at those correlations in Table 3 that are adjusted for occupation, we see that older respondents obtain higher scores on scale C (temperance), scale D (justice), scale E (faith) and scale F (charity). All these correlations may reflect genuine differences in these traits at various ages. The reasons for the significant correlation of scale F (charity) with gender such that women appear to be more open than men, is not known. Educational level correlates significantly with scale E (faith) and scale G (hope) such that a better education is associated with a decreased confidence in others and an increased ability to cope with change. Differences between ethnic groups, and between speakers of English as a second language (versus the rest), were examined for each Giotto scale using analysis of variance. No significant differences were found.

Table 2
The intercorrelations between the seven Giotto scales in the standardisation study. (Split-half reliabilities are shown in the diagonal) ($N=701$) and correlations with the Orpheus Lie scale ($N=504$)

	A	B	C	D	E	F	G	Lie
A	(0.72)	0.49**	0.04	0.07	−0.02	0.01	0.12	0.03
B		(0.71)	−0.34**	0.16**	−0.23**	−0.04	0.34**	0.15**
C			(0.75)	0.19**	0.48**	0.33**	−0.26**	−0.23**
D				(0.71)	0.01	0.43**	0.29**	−0.14**
E					(0.73)	0.40**	−0.38**	−0.18**
F						(0.75)	−0.10	0.31**
G							(0.76)	0.11*

* $p<0.05$.
** $p<0.01$.

Table 3
Correlations of Giotto scales with age, sex and educational level in the standardisation study

	Raw correlation			Pooled within occupations		
	sex	age	education	sex	age	education
A	−0.07	0.09*	−0.07	−0.01	0.06	0.04
B	0.00	0.07	−0.00	0.04	0.10	0.14*
C	−0.05	0.17**	−0.13**	−0.05	0.10*	−0.09
D	−0.02	0.08*	0.07	0.02	0.29**	0.26**
E	0.00	0.16**	−0.23**	0.03	0.10**	−0.22**
F	0.20**	0.28**	0.07	0.15**	0.30**	0.09
G	−0.07*	−0.07	0.13**	−0.05	0.00	0.20**
N	530	436	353			

Sex: male=1, female=2; education: from none=1 to Ph.D.=6. Age in years.

3.3. *Validity*

Content validity is demonstrated by the nature of the adjectives that show the most extreme loadings on each scale for both positive and negative directions. These are shown in Table 4.

The correlations between the seven Giotto scales and the seven Orpheus minor scales are shown in Table 5. It will be noted that, apart from the special case of scale F (charity), all correlations with the equivalent scales are significant, ranging from 0.40 for the correlation of scale A (prudence) with proficiency to 0.58 between scale C (temperance) with patience. Scale F (charity) has no direct equivalent in Orpheus. As disclosure is in effect a reverse lie scale, we can see that the correlations

Table 4
Content validity for Giotto, demonstrated by the adjectives obtaining high and low loadings ($N=701$)

Scale A (prudence)	
Positive	wise, competent, serious, cautious, sensible, proficient
Negative	absentminded, careless, reckless, peculiar, silly, foolish
Scale B (fortitude)	
Positive	determined, forceful, punctual, industrious, tireless, resolute
Negative	idle, untidy, irresponsible, unreliable, unkempt, whimsical
Scale C (temperance)	
Positive	patient, inoffensive, calm, gentle, self-controlled, non-violent
Negative	aggressive, angry, stormy, threatening, intimidating, bad-tempered
Scale D (justice)	
Positive	impartial, ethical, just, balanced, unbiased, facilitative
Negative	touchy, suspicious, resentful, hard-done-by, scape-goated, complaining
Scale E (faith)	
Positive	loyal, obedient, dutiful, trusting, courteous, modest
Negative	arrogant, egotistical, opinionated, vain, self-centred, conceited
Scale F (charity)	
Positive	compassionate, generous, benevolent, sincere, honest, unselfish
Negative	covetous, crafty, envious, spiteful, unappreciative, grasping
Scale G (hope)	
Positive	enthusiastic, enterprising, creative, imaginative, resourceful, inventive
Negative	disheartened, discouraged, weary, backward-looking, cynical, pessimistic

Table 5
The correlations of the Giotto scales with the Orpheus minor scales ($N=504$)

Orpheus scale	Giotto scale						
	A	B	C	D	E	F	G
P1: proficiency	**0.40****	0.23**	0.14**	0.09*	0.10*	−0.03	0.03
P2: work-orientation	0.29**	**0.56****	−0.35**	0.25**	−0.41**	−0.15**	0.47**
P3: patience	0.01	−0.27**	**0.58****	0.19**	0.38**	0.23**	−0.11*
P4: fair-mindedness	0.08	0.26**	−0.05**	**0.47****	−0.30**	0.14**	0.46**
P5: loyalty	0.09*	−0.27**	0.38**	−0.16**	**0.57****	0.18**	−0.49**
P6: disclosure	−0.23**	−0.18**	−0.18**	−0.17**	−0.24**	**−0.31****	−0.09**
P7: initiative	0.12**	0.42**	−0.28**	0.31**	−0.54**	−0.11**	**0.61****

Table 6
Correlations between the seven Giotto scales and the big five as measured by the five Orpheus major scales ($N=504$)

Giotto	Orpheus				
	F	A	C	E	D
Scale A (prudence)	0.01	0.17**	0.08	−0.19**	0.41**
Scale B (fortitude)	0.14**	0.47**	−0.17**	−0.26**	0.19**
Scale C (temperance)	−0.08	−0.50**	0.16**	−0.04	0.17**
Scale D (justice)	0.13**	0.07	−0.34**	−0.26**	−0.01
Scale E (faith)	−0.14**	−0.54**	0.40**	0.18**	0.37**
Scale F (charity)	0.07	−0.31**	−0.03	0.03	0.13**
Scale G (hope)	0.28**	0.36**	−0.43**	−0.43**	−0.18**

The Orpheus scales are F (fellowship), A (authority), C (conformity), E (emotion) and D (detail). *$p<0.05$.
** $>p<0.01$.

of the Giotto scales with social desirability are all relatively low, well below the target of 0.3 for all scales except scale F (charity) which is a quite acceptable −0.31.

Generally the patterns correlations between the Giotto scales and the Orpheus minor scales reflect the pattern of Giotto inter-correlations shown in Table 2. Table 6 shows the correlations between the Giotto scales and the Big Five personality traits as assessed by the Orpheus major scales.

Domain related validity is provided by cross correlations of Giotto scales with Orpheus items within the standardisation study. This is shown in Table 7.

The seven Giotto scales were validated against 12 supervisors' rating scales (see Table 8). Ratings ranged from below average through average, a little above average, much above average to exceptional. The ratings were carried out by the supervisors of 432 respondents in the standardisation study. The correlations in Table 8 are corrected for attenuation by dividing the correlation between the observed scores by the square root of the product of their reliabilities (Lord & Novick, 1968).

4. Discussion

The increased interest in integrity testing shown by personnel psychologists follows largely from a recognition that such tests address the attributes of candidates that are of most interest to prospective employers. Many of the more traditional personality tests, on the other hand, seem to relate to psychiatric diagnosis or to use terms more suited to student than employee populations. A further difficulty with traditional personality tests arises from a long-standing conviction that personality terms are and should be morally non-evaluative in nature. This belief can be traced back to the work of Allport (1937), who claimed that the so-called vices and virtues were more

Table 7
Correlations between Giotto scales and Orpheus items. Item/scale correlations appear on the right. The 3 highest and lowest correlations are reported for each scale. ($N=504$)

Scale A (prudence)	
0.31	I have a reputation for being good at checking detail carefully
0.26	Nobody has ever considered me to be unreliable
0.25	I never do anything without a good reason
−0.32	I am sometimes too rash in making decisions
−0.30	Some people might say that my working style is untidy (also fortitude 0.27)
−0.29	On some occasions I find it difficult to concentrate properly on what I am doing
Scale B (fortitude)	
0.33	I don't care if some people think I am pushy as long as I get things done
0.29	I am particularly effective at persuading others to do as I would like
0.26	I believe I may be more able to cope with disaster than most people
−0.31	I occasionally find it difficult to stand up to my seniors
−0.29	I expect some people think I am too 'laid back'
−0.28	I sometimes wish I was more able to speak my mind
Scale C (temperance)	
0.37	I am very patient with people even when I know they are wasting my time
0.25	When I feel angry with someone I almost always try not to let it show
0.24	I am always very calm and collected
−0.41	I expect some people say that I am often too harsh in my judgement of others
−0.40	In all honesty I would have to admit that I am often a bit bossy
−0.39	I am sometimes ruthless in order to succeed
Scale D (justice)	
0.26	I am sure that everyone I know thinks of me as a cool and rational decision maker
0.25	I am exceptionally good at calming everyone down when feelings are running high
0.20	People are usually honest with me
−0.34	There always seems to be at least one difficult person within every team
−0.26	Most people are unfortunately too stupid to realise which things in life are important
−0.26	I would never advise leaving an important job to a subordinate
Scale E (faith)	
0.38	People have sometimes told me that I am not forceful enough
0.36	I tend to feel uncomfortable if I go against the rules
0.33	I hate being forced to cut corners to get a job completed on time
−0.40	I am particularly skilled at asking penetrating questions
−0.31	I find clerical work somewhat tedious
−0.29	I find routine administration boring and prefer to leave it to others
Scale F (charity)	
0.29	If someone accidentally gave me too much change I would always tell them
0.25	I always tell the truth
0.24	I have never ever been dishonest
−0.37	I have no problem in twisting the truth when it's really necessary
−0.32	Making a good first impression is sometimes more important than being truthful
−0.31	It's essential to be aggressive sometimes in order to get your way

(continued on next page)

Table 7 *(continued)*

Scale G (hope)	
0.40	I always take a very optimistic view of my chances of success at work
0.34	I usually find it easy to enthuse others with my ideas
0.33	I find no difficulty whatsoever in getting people to take notice of me
−0.35	I often feel anxious that I might have made a wrong decision
−0.35	I can sometimes be at a loss for words when I meet people for the first time
−0.33	I find it much easier to work when there are clear rules to guide me

Table 8
Correlations between Giotto scales and 12 supervisors' ratings, corrected for attenuation. Ratings were: (1) level of self-confidence, (2) ability to generate new ideas, (3) ability to make tough decisions, (4) attention to detail, (5) breadth of vision, (6) time-keeping, (7) level of self-control, (8) suspicion of authority, (9) obedience to company policy, (10) trustworthiness, (11) ability to cope with change and (12) level of ambition ($N=432$)

	Giotto scale						
	A	B	C	D	E	F	G
1	0.18*	0.37**	−0.13	0.23**	−0.15	−0.07	0.31**
2	−0.07	0.13	−0.13	0.13	−0.23**	−0.07	0.23**
3	0.02	0.25**	−0.08	0.13	−0.12	0.03	0.23
4	0.03	0.13	0.11	0.23**	0.25**	0.18*	0.08
5	−0.07	0.13	−0.13	0.15	−0.18*	−0.13	0.19**
6	0.08	0.27**	0.10	0.20*	0.22**	0.15	0.05
7	−0.08	0.02	0.10	0.05	0.22**	0.07	−0.06
8	0.02	−0.05	−0.02	−0.17*	−0.18*	0.02	−0.13
9	0.08	0.03	0.16*	0.08	0.38**	0.20*	−0.11
10	−0.03	0.07	0.11	0.22**	0.25**	0.21**	0.02
11	0.00	0.20*	−0.05	0.05	−0.02	−0.07	0.21**
12	0.10	0.28**	−0.21**	0.00	−0.23**	−0.26**	0.26**

* Significance <0.05.
** Significance <0.01 (2−tailed).

relevant to ethics than psychology. The medical model implies that a personality profile is merely diagnostic (there are no 'right' or 'wrong' answers), but this can lack conviction where the profile is representative of proneness to depression or antisocial personality disorder and the respondent's career is dependent on the result. Of course, clinical personality traits can be avoided in occupational settings, using, for example, only the Big Five. However, it has been claimed that the trait of conscientiousness, which of the Big Five provides the strongest prediction of employment success (Ones, Schmidt, Viswesvaran, & Lykken, 1996; Sackett & Wanek, 1996) is more appropriately seen as an integrity trait. For an employer, conscientiousness is normally a virtue.

Thus, it appears that in spite of a widespread desire by psychologists not to make morally

evaluative judgements concerning their clients, it may well be those traits for which this is most questionable that are most important in personnel settings. The Allport and Odbert (1936) deliberate exclusion of vices and virtues from their list of potential items carried over into other factor analytic studies, such as that used by Cattell (1943) in developing the 16PF. Similarly, the original studies by Norman and by Goldberg were also based on the Allport and Odbert list, so that the Big Five model itself is a product of this exclusion. What would have happened had such exclusions not occurred? Interestingly, Almagor, Tellegen, and Waller (1995) report a study of a factor analysis of natural language trait descriptors where evaluative adjectives are retained rather than eliminated and conclude that the data better fits a seven factor rather than a five factor model. Five of their seven factors are, fairly closely related to the Big Five, while the two new factors reflect positive and negative evaluation respectively. Nonetheless, quite a number of items do migrate from the Big Five scales of agreeableness and neuroticism to the new factors, while the inter-factor structure itself is complex.

The seven-factor model identified by Almagor is not, however, the same as that used in Giotto, nor does it map on to the Prudentius model in the same way. Consequently evidence for the validity of this model must be found elsewhere. An alternative might be the 'wisdom of the ancients' approach taken by Eysenck (1947) in basing his two-factor model on Galen's theory of temperament related to the four humors (phlegmatic, sanguine, choleric and melancholic). Galen, like Prudentius, lived in the later days of the Roman Empire and generated a theory that has continued to have influence to this day. Unlike Galen's typology, that of Prudentius is quite clearly judgmental. Character is not seen as a product of nature but as the result of choices, freely made. Bad choices are habit forming and become ingrained, preventing a person from becoming fully rational. However, while Galen's deterministic philosophy has found its echoes in behaviourism, so we can recognise themes from Prudentius's striving for the rational in modern cognitive approaches to therapy. Furthermore, the seven bipolarities suggested by the model of Prudentius form a very good fit to existing concerns in integrity testing. Additionally, the model is able to take full account of the moral dimension to behaviour that appears to be central to any assessment of integrity.

All personality questionnaires have a problem when attempting to assess socially desirable behaviour and this is particularly true of integrity tests. Indeed, if we treat social desirability and lying as synonymous then paradoxically many integrity tests have a negative correlation with honesty (Rust, 1975; Sackett et al., 1989; Guastello & Rieke, 1991). However, as Ones et al. (1993) point out, response distortion to the extent that it exists does not seem to destroy the criterion related validity of integrity tests. With the Giotto questionnaire, response distortion of this type is minimised by the use of an ipsative framework. Response adjectives have been paired in such a way that a social desirable response is not possible, rather the respondent must choose between attributes that are desirable or undesirable in different ways. Furthermore, by randomly assigning the order of the items, acquiescence effects are also reduced to a minimum. The success of this overall strategy is demonstrated in Table 2. The largest correlation of any Giotto scale with social desirability is 0.31, and most are considerably less than this.

The ultimate test for the usefulness of any integrity test is its validity. Evidence for this is drawn from a number of different sources, demonstrating predictive, construct and concurrent validity. Some evidence for criterion related validity is obtained from the correlations with supervisors' ratings. These show some support, particularly for scale B (fortitude), scale E (faith), scale F

(charity) and scale G (hope), with some weaker support for scale D (justice). Results for scale A (prudence) and scale C (temperance) are less encouraging. However, this evidence needs to be considered in the light of the support from other forms of validity and the known unreliability of supervisors' ratings. Further supportive studies for criterion related validity is required.

The application of neural network technology to psychometrics is a new development only made possible by advances in software over recent years. Network programs themselves are primarily actuarial devices that maximise prediction and while this is clearly a desirable quality in psychometric tests, existing conventions require a higher level of understanding of the mechanisms involved. However, the use of neural technology in Giotto is very conservative, being limited to a testing of assumptions and the provision of extra item statistics for guidance in item selection. Giotto provides a novel framework for the investigation of a broad spectrum of integrity traits and it is to be hoped that further work in this area will establish a useful link between traditional social values and the assessment of integrity.

References

Allport, G. W. & Odbert, H. S. (1936). Trait names: a lexical study. *Psychological Monographs*, 47 (1, Whole No. 211).

Allport, G. W. (1937). *Personality: a psychological interpretation*. New York: Holt.

Almagor, M., Tellegen, A., & Waller, N. G. (1995). The Big Seven model: a cross-cultural replication and further exploration of the basic dimensions of natural language trait descriptors. *Journal of Personality and Social Psychology, 69*, 300–307.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: a meta-analysis. *Personnel Psychology, 44*(1), 1–26.

Bergman, J. (1926). Aurelii Prudentii Clementis. *Corpus Scriptorum Ecceliasticorum Latinorum* (Vol. 61). Vienna: Hoelder-Pichler-Tempsky A.G.

Camara, W. J., & Schneider, D. L. (1994). Integrity tests: facts and unresolved issues. *American Psychologist, 49*, 112–119.

Cattell, R. B. (1943). The description of personality: basic traits resolved into clusters. *Journal of Abnormal and Social Psychology, 38*, 476–506.

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. (Eds.) (1989). *Tenth mental measurement yearbook*. Buros Institute of Mental Measurement, University of Nebraska Press, Lincoln, NE.

DTI (1994). *Neural Computing: best practice guidelines for developing neural computing applications*. London: Department of Trade and Industry, UK Government.

Eysenck, H. J. (1947). *Dimensions of personality*. London: Routledge and Kegan Paul.

Galton, F. (1884). Measurement of character. *Fortnightly Review*, 42.

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.

Guastello, S. J., & Rieke, M. L. (1991). A review and critique of honesty test research. *Behavioral Sciences and the Law, 9*, 501–523.

Haynes, S. N., Richard, D. C., & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Neuralware (1995). *Neural explorer handbook*. Seattle: Neuralware.

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.

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

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*. London: The Psychological Corporation.

Rust, J. (1996). *Orpheus handbook*. London and San Antonio: The Psychological Corporation.

Rust, J. (1997). *Giotto handbook*. London and San Antonio: The Psychological Corporation.

Rust, J. & Golombok, S. (1999). *Modern psychometrics* (2nd ed.). London: Routledge.

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

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness and reliability for personnel selection. *Personnel Psychology, 49*(4), 787–830.

Sayers, D. (1955). *Dante's Purgatorio* (English translation). London: Penguin books.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43*, 627–670.

Schmidt, M. J., & Ryan, A. (1993). The Big Five in personnel selection: factor structure in applicant and non-applicant populations. *Journal of Applied Psychology, 78*, 966–974.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel Psychology, 44*(4), 703–742.

Thompson, H. J. (1949). *Prudentius* (Latin with an English translation). Cambridge, MA: Harvard University Press.

US Congress, Office of Technology Assessment (1990). *The use of integrity tests for pre-employment screening* (OTA-SET-442). Washington, DC: US Government Printing Office.