Integrity Tests and the Five-Factor Model of Personality: A Review and Empirical Test of Two Alternative Positions

Bernd Marcus[*]
University of Western Ontario

Stefan Höft
German Aerospace Center

Michaela Riediger
Max Planck Institute for Human Development

The psychological meaning of integrity test scores has been explored predominantly in relation to the five-factor model of personality (FFM). Two alternative positions on this topic can be identified in the literature which state, respectively, that integrity tests measure (a) a higher-order factor of personality covering three FFM dimensions or (b) a linear composite of numerous facets from various domains within the FFM. An empirical test of these alternative positions, using structural equation modeling, revealed that the value of both views depended on the type of integrity test examined. With a personality-based integrity test, position (a) had to be refuted, whereas position (b) was strongly supported. There was also more supportive evidence for position (b) with an overt test, but the difference was far less pronounced than for the personality-based measure. Possible consequences for theories on the role of personality in personnel selection are discussed.

The criterion-related validity of integrity tests for personnel selection has been established through extensive research, as summarized in a large-scale meta-analysis (Ones, Viswesvaran, & Schmidt, 1993). There is also evidence that integrity tests are among the most effective selection tools in terms of incremental validity over cognitive abilities (Schmidt & Hunter, 1998). Hence, the practical value of these instruments in general seems no longer a matter of debate. Consequently, the major research interest shifted to more theoretical issues in recent years (compare the relative space devoted to construct- and criterion-related validity in two major, successive reviews of the field by Sackett, Burris, & Callahan, 1989; Sackett & Wanek, 1996). The crucial point now appears to be why integrity tests are valid, which is closely connected to the

question what they measure. This issue has important theoretical and practical implications. For instance, in a discussion on the appropriate level of personality assessment for personnel selection, both proponents of a "broad" (Ones & Viswesvaran, 1996) and "narrow" (Schneider, Hough, & Dunnette, 1996) measurement approach have tried to bolster their contradictory positions by findings that instruments like integrity tests tend to predict overall job performance with higher validity than standard personality inventories (cf. Barrick, Mount, & Judge, 2001). Obviously, the more valid a test is, the more important is its underlying construct − or constructs, when it is heterogeneous − as a determinant of occupational success. Hence, to know what valid tests measure can have significant impact on theories of job performance.

Of course, it would be desirable to begin an article on the construct measured by integrity with a definition of integrity. Unfortunately, this would have been at odds with the way these tests are developed and the objective of the present article. Although there are scholarly discussions on

*Address for correspondence: Bernd Marcus, Department of Psychology & Bachelor of Administrative and Commercial Studies Program, University of Western Ontario, Social Sciences Centre, London, ON, Canada. E-mail: bmarcus3@uwo.ca

the nature of integrity (e.g., Becker, 1998; McFall, 1987), none of the more established integrity tests was based on any of these definitions. In fact, the label "integrity" was attached to these tests long after the most prominent exemplars were already in use, replacing earlier labels like "honesty" (cf. Sackett & Wanek, 1996). As the present article is concerned with the constructs actually measured by the tests now referred to as integrity tests, not with the meaning of the term later chosen to label them, we prefer to avoid any presuppositions on the meaning of those constructs.

Sackett *et al.*'s (1989) distinction between "overt" and "personality-based" integrity tests is now a commonplace framework for research on this topic. Roughly speaking, overt tests contain relatively transparent items directly related to counter-productive behavior (e.g., "Have you ever thought of stealing money from your workplace without doing it in reality?"). Personality-based tests are composed of items often adopted from traditional personality inventories, the relationship of which to the criterion is not always obvious but empirically supported (e.g., "I am more sensible than adventurous.").

Although other theoretical approaches have been occasionally investigated (see Sackett & Wanek, 1996, for a more comprehensive review), the vast majority of empirical research on the construct(s) measured by integrity tests can be organized within the descriptive framework of the five-factor model of personality (FFM, or "big five"; e.g., Wiggins, 1996). In this paper, the two most influential views on integrity and the FFM are reviewed with respect to their theoretical, empirical, and methodological foundations. In addition, an empirical study is conducted that provides the first attempt to test both hypotheses with the same data set. Results of this study are presented and discussed subsequent to the review section.

Two Views on Integrity and the FFM

Before proceeding with a detailed verbal review of both positions, a graphic representation may serve as an overview of this section. Figure 1 provides a (somewhat simplified) visualization of the way integrity is conceptualized within the FFM framework, according to the respective views. Both subfigures contain identical graphs depicting the FFM along with two additional hierarchical levels of personality structure: on the left, the highest-order, or *superfactor level* above the big five; in the middle, the actual five-factor, or *dimension level*; and to the right of the latter, a lower-order, or *facet level* of more narrow constructs not further specified. The appropriate place of integrity within this taxonomy, as hypothesized in the respective positions, is indicated by the darkened areas in both subfigures. As outlined in detail below, position (b) differs from position (a) in at least two ways: Integrity is *not* seen as a homogenous latent construct (indicated by
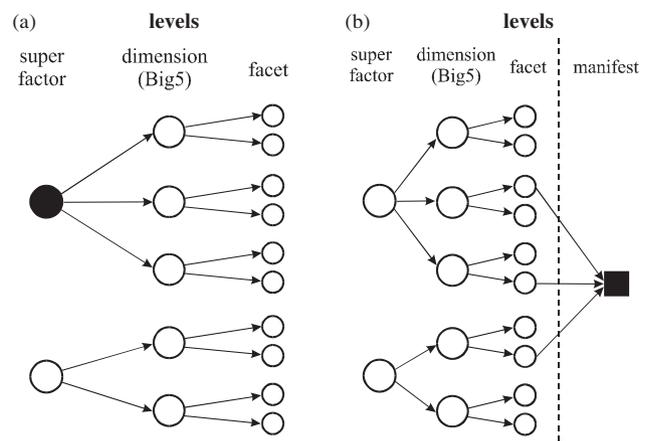
Figure 1. Simplified visualization of the competing views on integrity and the five-factor model of personality.

presenting it as a box rather than a circle), and there is no particular place where to locate it within the FFM (indicated by the dotted line in the diagram).

(a) The Superfactor Hypothesis: Integrity as a Higher-Order Factor of Personality

Ones, Schmidt, and Viswesvaran (1994a) hypothesized that "... [integrity tests] tap into a general broad personality trait, much broader than any one of the big five. Integrity tests tap into a higher-order factor that includes Agreeableness, and Emotional Stability, as well as Conscientiousness" (p. 28). If this statement were true, it would mean that there is a superfactor of personality above the big five level and that it is this very dimension that is at the core of personal determinants of overall job performance beyond general intelligence.[1] It is somewhat puzzling to note that the bases for such a far-reaching conclusion have only received limited attention in the following years.

Ones and colleagues amassed an enormous volume of both primary (1994b) and secondary (1994a) data to support this position. Put briefly, they formed composites of different integrity tests (using entire scales as elements), either belonging to the overt or personality based type (cf. Sackett *et al.*, 1989), then did the same for different scales held to measure the three relevant FFM dimensions, and calculated correlations between their integrity and personality composites. In the next step, they collapsed both kinds of integrity tests and/or all three big five dimensions into one construct, respectively, and again computed the composite correlations. At the final level of "all integrity" and "all personality" aggregates, they arrived at almost perfect correlations between both composites. This was taken as evidence that integrity tests measure a superfactor of personality. Although the sample sizes involved are impressive, this extremely condensed

way of data integration merits some methodological as well as conceptual remarks.

To understand the implications of this extensive use of composite correlations, an inspection of the respective formula is informative (Ghiselli, Campbell, & Zedeck, 1981, p. 181):

$$r_{c_{z_x}c_{z_y}} = \frac{\bar{r}_{x_k y_m}}{\sqrt{\frac{1}{k} + \frac{(k-1)}{k}\bar{r}_{xx'}}\sqrt{\frac{1}{m} + \frac{(m-1)}{m}\bar{r}_{yy'}}} \qquad (1)$$

with c_{z_x}, composite of the standardized predictor set $c_x = x_1 + x_2 + \ldots + x_k$; c_{z_y}, composite of the standardized criterion set $c_y = y_1 + y_2 + \ldots + y_m$; k (m), number of components in the x- $(y$-) composite; $\bar{r}_{x_k y_m}$, average of all bivariate correlations between x- and y-components; $\bar{r}_{xx'} (\bar{r}_{yy'})$, average intercorrelation between pairs of components in the x- $(y$-)composite.

The denominator of (1) corrects for the incongruence between different measures of the same construct (imperfect within-composite correlations), assuming that they in fact tap into one construct. According to Ghiselli *et al.*, one major consequence is: "The lower the intercorrelations among the components (the less homogeneous the components), the higher the correlations between the two composites" (p. 175). Lower within-composite correlations decrease the value of the denominator. As this value is always smaller than one, decreasing it causes the result of the equation to rise.

That mechanism is equivalent to the way traditional attenuation correction for low reliability leads to increased correlations. However, whereas imperfect reliability is usually regarded as a statistical artifact, low within-composite correlations might well be taken as evidence that the elements involved actually measure different constructs. A substantive interpretation of this composite as a homogeneous construct may therefore not always be justified. Moreover, the effect of within composite heterogeneity is further amplified as the number of aggregated elements (indexed by k and m in the formula) increases. Thus, there is good reason to assure the theoretical meaningfulness of a composite on every single level of aggregation before drawing firm conclusions on the values found with composite correlations.

There are at least three levels of aggregation involved in the analyses just decribed. First, Ones *et al.* began their analyses at the level of entire scales, which are themselves aggregates. As long as it is not established that these single tests measure a one-dimensional construct, at least at a higher-order level, it would be difficult to interpret the meaning of that construct across different instruments. While Ones *et al.* did not address this issue, empirical evidence on the internal structure of integrity tests stemming from other sources is not unequivocal. Exploratory factor analyses at the item level have consistently revealed several dimensions (e.g., Cunningham & Ash, 1988; Harris, 1987; Harris & Sackett, 1987; Kochkin,

1987; Paajanen, 1988; Wanek, 1995). However, most of these studies have also provided evidence supporting homogeneity by findings of a marked drop in eigenvalues after the first factor (Cunningham & Ash, 1988; Harris, 1987; Harris & Sackett, 1987), or by confirmation of a Rasch model, which assumes unidimensionality (Harris & Sackett, 1987). The findings reviewed so far refer to the overt type of integrity tests only. For at least one personality-based integrity test, eigenvalues showed a much more even trend in PCA (Paajanen, 1988). Moreover, the latter type of tests are either explicitly developed to measure several independent constructs (Borofsky, 1993), or are based on Gough's (1960) CPI-Socialization scale (Gough, 1971; Hogan & Hogan, 1989), which itself was repeatedly found to be multidimensional by rational (Gough & Peterson, 1952) and empirical (Collins & Bagozzi, 1999) examination (but see Rosen, 1977, who also confirmed a one-factorial solution). More recent studies involving item level analyses of both types of integrity tests simultaneously either yielded one-factor (Hogan & Brinkmeyer, 1997) or four-factor (Wanek, Sackett, & Ones, 2003) solutions.

While there appear to remain a number of unresolved issues concerning the dimensionality of single integrity tests, Ones and colleagues explicitly addressed their next two levels of aggregation where they first collapsed different instruments into composites for the two kinds of integrity tests, and then into one composite for the entire domain. Ones (1993) provided evidence by means of confirmatory factor analysis (CFA) that there are distinct factors by test type as well as a higher-order factor loading highly on both subdimensions. Thus, there is indeed evidence that different integrity tests do have something in common when scale scores are analyzed.

Before finally correlating their integrity composite with the three FFM dimensions composite, Ones *et al.* (1994a, b) applied the same procedure as above to the scales of numerous personality inventories. In this case, scales of presumably internally homogenous content but very diverse theoretical origins are sorted into composites, which are then named after the respective FFM terms. The overall pattern of relationships in both the primary and the meta-analytic study was characterized by relatively small differences between within- and between-composite correlations. This is an almost ideal correlational pattern for maximizing composite correlations, but it may well be an artifact of sorting scales into composites that do not optimally represent the constructs they are held to measure and at the same time possess partial content overlap (or secondary loadings) with constructs they presumably not measure.

Support for the assumption of a general personality factor above the FFM level comes from a more recent study by Digman (1997). He reanalyzed 14 studies with big five inventories and found evidence for two second-order dimensions apparent across all data sets. One of these

was comprised of emotional stability, agreeableness, and conscientiousness – the very factors Ones *et al.* collapsed at their highest level of aggregation – and labeled "α" by Digman. Digman's substantive interpretation of this super-factor was that it measures the general success of the socialization process. However, he also considered the alternative explanation that this abstract dimension might reflect a response set, social desirability. It is also noteworthy that Ones *et al.* based their aggregation of the three big five factors on mean observed correlations around .20, indicating rather modest relationships.

In summary, the superfactor hypothesis is an ambitious interpretation of the construct measured by integrity tests. However, the extensive use of composite correlations and the highly complex multi-step procedure of data integra-tion leave room for alternative interpretations. A more direct primary empirical investigation may help to clarify these issues.

(b) The Several-Facets Hypothesis: Integrity as a Composite of Single Facets from Various FFM Domains

In defense of narrow traits for personnel selection, (Schneider *et al.*, 1996; see also Hough & Schneider, 1996) made a point on integrity tests that introduced a position contradictory to the superfactor hypothesis: ". . . examination of some of the more prominent measures of integrity . . . reveals that they are actually constellations of narrow personality traits" (p. 644). Therefore, the authors do not assume that integrity tests measure one construct at some higher yet substantive level of aggregation. Rather, integrity tests are held to be comprised of a number of distinct, narrow traits (e.g., at the facet level below FFM dimensions), which need not be related among each other. This does not preclude potentially high correlations among tests at the scale level if they are all constructed in a similar manner.

Schneider *et al.* based their contention primarily on the historical roots of some integrity tests, which stand in the tradition of combining items or narrow personality traits found to differentiate between contrasted groups of offenders and non-offenders, without considering the homo-geneity of the entire scale for construction. By inspection, however, these criterion-keyed "prominent measures" all appear to belong to the category of personality-based tests (cf. Borofsky, 1992; Gough, 1971; Hogan & Hogan, 1989; Paajanen, 1988). Authors of leading overt integrity tests often preferred a rational scale construction strategy that emphasized attitudes directly linked to the behavior (e.g., Brooks & Arnold, 1989; Jones, 1991). Schneider *et al.* did not explicitly distinguish between different integrity test categories.

The paucity of empirical research that directly addressed relationships between integrity tests and the FFM at the facet level makes it difficult to assess the validity of the several-facets hypothesis. One of its basic assumptions is that integrity tests should be differentially related to the more narrow traits comprising each FFM dimension. The only study Schneider *et al.* cited in favor of their position was a relatively small-scale ($N = 123$) investigation (Costa & McCrae, 1995), where the Revised NEO-Personality Inventory (NEO-PI-R) facets in fact showed markedly different associations with the Hogan Reliability Index (cf. Hogan & Hogan, 1989) within all factors (similar patterns occurred for two integrity tests related to the Hogan Personality Inventory [Hogan, 1986] in a study by Murphy & Lee, 1994, and for another personality-based test related to the NEO-PI-R, Hakstian, Farrell, & Tweed, 2002). This stands in sharp contrast to the sample sizes of some 10,000 subjects reported by Ones *et al.* (1994a) for their meta-analyses, as well as to the $N = 1365$ in their primary investigation (1994b). Moreover, Ones' research strategy is in no way comparable with the studies cited in favor of the several-facets hypothesis.

Objectives of the Present Study

Hypotheses (a) and (b) propose different explanations for how integrity relates to the normal personality sphere but can neither be rejected nor verified on the basis of present evidence. The study to be presented below is meant to address this controversy in a more comparative manner than earlier investigations. Because of conceptual incon-gruence, however, an empirical comparison of both positions could be approached only indirectly, as will be shown next.

A Note on the Conceptual Difficulties in Testing Both Positions, and a Possible Solution

Position (a), which was labeled the "superfactor hypoth-esis," postulates a hierarchically structured factorial model where integrity tests of both kinds are loaded by a highly general factor of personality. This very abstract dimension is comprised of the three factors of conscientiousness (C), agreeableness (A), and emotional stability (ES) at the next lower level (in that order, Ones, 1993). While they were very explicit with respect to the big five, Ones *et al.* (1994a, b) did not address the next lower level of facets. Thus, it may be concluded that it was the dimensions as defined by the FFM that count, according to this view, not just single facets within dimensions.

The theoretical image of the integrity construct behind this view clearly corresponds to what different authors, taking different perspectives, alternatively labeled a "multi-faceted trait" (Schneider & Hough, 1995), "effect indicator model" (Bollen & Lennox, 1991), or "latent model" (Law, Wong, & Mobley, 1998). It is not at issue that testing such a model would require a factor-analytic approach and, if the state of theory development allows for, CFA may be most

appropriate. More specifically, in the superfactor view of integrity, a general factor of personality ("α" or socialization according to Digman, 1997) loads on three FFM dimensions as well as on both integrity subfactors, which means that α and general integrity can be taken as equivalents. There is no lower-order structure specified in this view, that is, integrity tests may or may not possess a complex internal structure but this does not affect the general assumption that they all tap into one big construct. It follows that every single integrity test is held to measure one general factor at the scale level.

This implies several distinct hypotheses, which can be tested by a three-step CFA procedure. First, a higher-order factor may be modeled and fitted for each type of integrity test separately. Given that these models hold, the next step would be to specify a general factor that loads on both integrity test categories (Ones *et al.* addressed this second step but omitted the first). Finally, it has to be demonstrated that the data can be fitted to a model in which the same integrity factor loads on C, A, and ES but not on the remaining two FFM dimensions.

The procedure just described may be appropriate to test the superfactor hypothesis but is not very informative about the merits of the several-facets position. A general, although perhaps weak factor may drive the relationships between different integrity tests and the FFM, while a significant amount of variance unexplained by this factor may be attributable to unique facets within integrity tests and, accordingly, to unique FFM facets. Of course, there is a trade-off between both positions: The more powerful a general factor is as an explanatory construct, the less weight may be given to the specific facets. But the major difficulty of testing both positions competitively stems from the fact that both are conceptually different rather than directly contradictory.

The several-facets-hypothesis assumes that integrity tests are composed of a number of narrow personality traits that are themselves homogeneous but are not necessarily related among each other via one higher-order latent construct. Instead, they are aggregated solely because all are predictive of one or several criterion constructs (theft, counter-productivity, job performance), the internal structure of which is largely unknown and presumably complex. That is, integrity tests are actually not conceptualized as single tests but as test batteries; their scores are weighted linear composites of several predictors. In the theoretical conceptions cited above, this idea refers to the terms "compound traits" (Hough & Schneider, 1996), "causal indicator model" (Bollen & Lennox, 1991), or "aggregate model" (Law *et al.*, 1998). The corresponding statistical method is not factor analysis but multiple regression. Thus, an appropriate research strategy to test the several-facets hypothesis in isolation may look as follows: Select a number of relevant personality facets from FFM dimensions and regress integrity test scores as criteria on them. The sample of facets should cover several FFM

dimensions and should also incorporate differences between facets within domains, perhaps simplest by omitting some of the elements. If properly selected, the FFM facets should comprise a parsimonious set of predictors that accounts for the bulk of variance of integrity test scores.

Unfortunately, this is again supposed to be an appropriate way to test one hypothesis but not the other. We are faced with the dilemma that the different theoretical conceptualizations lead to incompatible methodological implications: a hierarchical factor structure modeled by CFA to test hypothesis (a); a linear predictor combination as in a multiple regression model to test hypothesis (b).

In the present study, this problem is approached by a two-stage strategy of hypothesis testing. At the first stage, the emphasis is placed on the superfactor hypothesis. Initially, a one-factorial CFA model is specified for both an overt and a personality-based integrity test and tested in competition to a model with correlated subfactors. If the one-factorial models fit to an acceptable extent, the FFM dimensions ES, A, and C are incorporated in a further step. In order not to change its psychological meaning, loadings of the general factor on its integrity facets are fixed to the values found in the preceding analyses. The same factor is allowed to load freely on the big five components. This provides a full test of the superfactor hypothesis (a) but is independent of the alternative position.

The several-facets hypothesis (b) is emphasized in the next phase of the investigation. Here, a series of multiple regression analyses is conducted in which both integrity tests are regressed on selected NEO facets using structural equation modeling (SEM) software (LISREL 8; Jöreskog & Sörbom, 1993). This permits, in addition to more traditional measures of variance accounted for (R^2, ΔR^2, adjusted R^2), the restriction of parameters and examination of model fit.

As a first step in these regression analyses, appropriate predictor facets of FFM dimensions in the version by Costa and McCrae (1992) were selected for both types of integrity tests. For the personality-based category, it was possible to base choices on findings from previous studies with NEO facets and integrity tests (e.g., Costa & McCrae, 1995), NEO and the CPI-So scale (McCrae, Costa, & Piedmont, 1993), factor-analytic studies yielding NEO-like components (Paajanen, 1988), test descriptions (e.g., Borofsky, 1993; Hogan & Hogan, 1989), and the rich literature on personality traits and deviant behavior (e.g., Hogan & Jones, 1983; Wilson & Herrnstein, 1985). Ten facets belonging to four NEO domains were selected (see Table 1) on this basis. Most of these lower-level constructs are associated with N, A, and C, respectively, but also half of the facets belonging to N, A, and C, are omitted (fixed to 0) in these analyses.

There was no equally solid ground to select facets for overt integrity tests. Here, it was decided to alter the sample of facets selected for personality-based tests, when the more attitudinal nature of overt instruments and their generally

Table 1. Model specifications for multiple regression analyses

Personality-based integrity test as criterion	Overt integrity test as criterion
Saturated model (30 facets as predictors, no restrictions)	Saturated model (30 facets as predictors, no restrictions)
1.1a: Several-facets/free parameters (N2, N5, E5, A2, A3, A4, C2, C3, C5, C6; rest fixed to zero)	2.1a: Several-facets/free parameters (N2, N5, E5, O6, A1, A2, C3, C6; rest fixed to zero)
1.2a: Superfactor/free parameters (all 18 facets from N, A, C; rest fixed to zero)	2.2a: Superfactor/free parameters (all 18 facets from N, A, C; rest fixed to zero)
1.1b: Several-facets/constrained parameters (in addition to 1.1a: β weights within same FFM domain constrained to equality)	2.1b: Several-facets/constrained parameters (in addition to 2.1a: β weights within same FFM domain constrained to equality)
1.2b: Superfactor/constrained parameters (in addition to 1.2a: β weights within same FFM domain constrained to equality)	2.2b: Superfactor/constrained parameters (in addition to 2.2a: β weights within same FFM domain constrained to equality)

Note: For NEO-PI-R facet labels, see Table 2.
FFM, five-factor-model of personality; NEO-PI-R, Revised NEO-Personality Inventory.

lower correlations with FFM *dimensions* (Marcus, Funke, & Schuler, 1997; Ones, 1993) suggested such changes. We dropped the NEO facets of altruism (A3), compliance (A4), order (C2), and self-discipline (C5), because these appeared to be among the least attitudinal aspects of A and C, which may explain the lower correlations of overt integrity with these FFM domains. Instead, the attitudinal dispositions of openness to values (O6), and trust (A1) were added to the list. Trust is, by definition, an attitudinal concept related to parts of the typical content of overt tests, while low openness to values most closely resembles the rigid attitudes some have speculated to be measured by this type of integrity tests (cf. Marcus, 2000, for a detailed review). A total of eight facets belonging to all five factors (cf. Table 2 for all NEO scale labels) were selected. Thus, only one third of the 18 facets belonging to N, A, and C plus two from the other domains are assumed to account for the variance in overt integrity test scores.

While these facets are selected for an isolated examination of the several-facets hypothesis, simultaneous theory testing required that the superfactor hypothesis be examined in a comparable way. For this purpose, additional multiple regression analyses are performed using all 18 facets of the three relevant factors postulated in hypothesis (b) as predictors.[2] This is not a completely adequate test of the respective hypothesis but could provide valuable insights in comparison with the alternative position. If the superfactor hypothesis holds, a linear combination of the facets representing the full variance of the three relevant dimensions should compare favorably to the more parsimonious selection of only eight or 10 facets in predicting integrity test scores.

In addition to the models described above, more parsimonious models are also tested. There, individual β weights within each factor are constrained to equality for

both positions tested. There are no conceptual grounds for this specification in the several-facets hypothesis but, according to the superfactor hypothesis, primarily FFM dimensions count. Thus, differences within those dimensions should not severely affect model fit. One final clue for the relative fit of specifications may be given by the saturated model (all 30 NEO facets predict integrity), which represents the optimum of explained variance in the criteria. Table 1 gives an overview of all multiple regression models tested at the second stage of this investigation. Taken together, we applied a simple principle in trying to solve the conceptual problems: If there is no golden rule to test both positions in one step, approach them from different perspectives. If the results converge, this provides tentative evidence that the findings are not affected by methodological differences.

Method

Sample and Procedure

Participants in this study were $N = 213$ undergraduate students at a German university majoring in diverse subjects (mainly economics, business administration, biology, and agriculture). Our choice of a student sample was based on the aim to investigate true relationships between trait constructs. Field settings, particularly with applicant samples, have often been shown to inflate such correlations and alter the factorial structure of personality tests (e.g., Brown & Barrett, 1999; Collins & Gleaves, 1998; Schmit & Ryan, 1993), whereas criterion-related validity remains largely unaffected by setting (e.g., Ones *et al.*, 1993; see measures section for evidence of the same findings with the present integrity measures). Because we were interested in uncovering a pattern of relationships at the construct level, we tried to avoid a setting where

Table 2. Means, standard deviations, Cronbach's α, and intercorrelations of study variables

Variable	M	SD	α	Overt integrity test					Personality-based integrity test					
				All	GT	PN	RA	I/P	All	MA	TA	PSC	RE	SS
NEO-PI-R dimension														
N: Neuroticism	86.7	23.4	.93	−.35	−.29	−.22	−.37	−.18	−.53	.25	.07	−.82	−.29	.08
N1: Anxiety	15.2	5.3	.80	−.25	−.20	−.20	−.26	−.11	−.30	.24	.06	−.61	−.08	.15
N2: Angry hostility	13.4	5.2	.78	−.30	−.30	−.20	−.24	−.16	−.61	.00	−.18	−.79	−.25	.07
N3: Depression	13.1	6.0	.86	−.26	−.26	−.10	−.31	−.09	−.34	.30	.06	−.61	−.20	.11
N4: Self-consciousness	16.3	4.3	.67	−.18	−.17	−.15	−.23	.01	−.21	.35	.33	−.57	−.14	.14
N5: Impulsiveness	16.5	4.8	.70	−.31	−.09	−.21	−.29	−.33	−.60	−.09	−.09	−.50	−.44	−.30
N6: Vulnerability	12.3	4.9	.83	−.31	−.27	−.17	−.36	−.13	−.37	.31	.18	−.71	−.25	.16
E: Extraversion	114.9	19.5	.90	.00	.14	−.10	.08	−.14	−.12	−.63	−.24	.27	.10	−.44
E1: Warmth	21.9	4.0	.73	.16	.28	.02	.14	.04	.10	−.28	−.11	.32	.12	−.22
E2: Gregariousness	19.4	4.8	.76	.01	.10	−.04	.05	−.08	−.08	−.42	−.06	.17	.00	−.28
E3: Assertiveness	17.1	5.6	.86	.05	.08	−.03	.14	−.07	−.04	−.70	−.40	.31	.20	−.17
E4: Activity	18.4	4.2	.73	.03	−.01	−.05	.14	−.05	−.11	−.51	−.23	.13	.16	−.29
E5: Excitement seeking	16.1	4.6	.56	−.34	−.11	−.18	−.25	−.44	−.39	−.34	−.03	−.10	−.19	−.63
E6: Positive emotions	22.1	5.3	.81	.11	.22	−.10	.12	.04	.03	−.30	−.10	.25	.09	−.26
O: Openness to Experience	120.9	18.5	.88	−.08	.09	−.15	−.04	−.17	−.11	−.02	−.17	.10	−.09	−.33
O1: Fantasy	20.1	5.3	.82	−.20	−.04	−.14	−.12	−.29	−.36	.01	−.21	−.21	−.34	−.17
O2: Aesthetics	20.2	5.6	.80	−.03	.02	−.11	−.06	.05	.08	.15	−.04	.10	.06	−.12
O3: Feelings	22.4	4.0	.71	−.13	−.00	−.24	−.12	−.07	−.11	−.03	−.16	−.05	.05	−.23
O4: Actions	18.5	4.3	.70	.04	.17	−.04	.09	−.10	.01	−.07	−.05	.29	−.03	−.47
O5: Ideas	18.7	4.7	.76	.04	.08	.04	.11	−.09	.03	−.19	−.17	.25	.01	−.17
O6: Values	21.0	3.4	.47	−.05	.20	−.13	−.04	−.19	−.08	.01	−.07	.07	−.10	−.25
A: Agreeableness	112.2	15.6	.86	.46	.47	.35	.29	.31	.38	.39	.30	.34	.05	−.04
A1: Trust	19.1	4.4	.78	.54	.78	.36	.39	.16	.22	.00	−.04	.37	.03	−.04
A2: Straightforwardness	18.0	4.0	.61	.41	.26	.32	.34	.32	.24	.55	.13	.04	.05	.14
A3: Altruism	21.3	3.8	.71	.21	.25	.09	.15	.17	.34	.07	.26	.42	.14	−.17
A4: Compliance	16.9	4.3	.64	.27	.23	.23	.13	.26	.50	.23	.54	.47	.06	.00
A5: Modesty	16.4	4.5	.75	.19	.07	.25	.07	.21	.05	.42	.14	−.10	−.06	.02
A6: Tender-mindedness	20.5	3.3	.55	.09	.22	.06	−.01	.03	.11	.19	.14	.10	.00	−.10
C: Conscientiousness	115.1	21.1	.92	.29	.05	.16	.27	.36	.65	−.16	.01	.46	.84	.24
C1: Competence	20.5	4.0	.72	.23	.17	.05	.24	.19	.43	−.34	−.09	.48	.54	.06
C2: Order	17.7	5.2	.76	.15	−.06	.06	.14	.26	.42	−.02	.06	.13	.68	.22
C3: Dutifulness	21.9	3.8	.68	.30	.13	.14	.29	.33	.59	−.02	.03	.45	.64	.21

Table 2. Continued

Variable	M	SD	α	Overt integrity test					Personality-based integrity test					
				All	GT	PN	RA	I/P	All	MA	TA	PSC	RE	SS
C4: Achievement striving	19.2	4.5	.71	.17	−.02	.12	.20	.18	.35	−.30	−.16	.33	.60	.07
C5: Self-discipline	18.9	5.1	.82	.28	.08	.16	.26	.32	.56	−.18	.02	.48	.73	.10
C6: Deliberation	16.8	5.0	.79	.23	−.03	.21	.16	.33	.62	.09	.15	.30	.67	.40
Overt integrity test:														
All scales combined	*199.8*	*23.6*	*.91*	—										
General trust (GT)	54.8	7.3	.81		—	.52	.63	.28		−.02	−.07	.42	.08	−.06
Perceived counter-productivity norms (PN)*	27.0	5.8	.76		.41	—	.65	.61		.12	.06	.36	.19	.25
Rationalizations (RA)*	65.8	9.3	.85		.52	.53	—	.57		−.03	−.04	.46	.24	.10
Behavioral intentions/phantasies (I/P)*	52.0	8.7	.80		.23	.48	.48	—		.26	.19	.36	.40	.44
Person-based integrity test:														
All scales combined	*175.2*	*17.1*	*.83*	.47	.26	.31	.35	.50	—					
Manipulativeness (MA)*	21.2	3.9	.65	.12	.04	.10	−.02	.24		—	.11	−.27	−.08	.32
Trouble avoidance (TA)	20.4	4.1	.67	.03	−.04	.01	−.06	.16		.19	—	.06	.04	.12
Positive self-concept (PSC)	57.0	9.6	.86	.45	.37	.29	.41	.29		−.13	.06	—	.47	−.07
Reliability (RE)	50.5	6.6	.79	.26	.08	.15	.20	.33		−.02	.05	.38	—	.28
Stimulus seeking (SS)*	20.1	4.5	.75	.22	.01	.17	.12	.35		.26	.09	−.03	.27	—

Notes: $N = 213$; correlations with an absolute value of .14 or higher are significant at the .05 level, two-tailed. Rows containing values for composite scales are printed in italics. All integrity scales are scored in the desirable direction (reverse-scored scales are indexed by an asterisk). In the lower portion of the table, zero-order correlations between raw integrity scale scores are given below the diagonal. Above the diagonal, coefficients for latent factors, as estimated in covariance structure analysis, are presented.

participants could be motivated to present themselves in a favorable light.

90.4% of the sample had at least 1 month of prior job experience and 54.8% reported an employment record of more than 1 year. One hundred and twenty-four (58.2%) participants were men. Mean age was 23.7 years with a range from 20 to 41 and a SD of 2.9. All participants were asked to volunteer for a research project on the campus site where they had to complete an extensive battery of psychological tests, including a number of instruments in addition to those described below (cf. Marcus, 2000). The study was conducted in groups of about 10–12 persons, each under supervision of a test administrator. The procedure took approximately 90–120 min for each group. Participants were paid an amount of DM 30 – (about 20 \$ U.S. at the time the study was conducted) as a compensation (see Ullman & Newcomb, 1998, for evidence that paying money reduces volunteer bias in research on deviant behavior).

Measures

Five-Factor Model. The NEO-PI-R (Costa & McCrae, 1992) was used in a German adaptation (Ostendorf & Angleitner, 2003; coefficients of congruence between the U.S. and the German version range from .96 to .98) as the FFM measure. The NEO-PI-R contains 240 Likert-type items in the score, with 48 measuring each of the five factors, neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Every higher-order factor, or domain, consists of six lower-order dimensions, or facets. Each facet is measured by eight items. These facets carry individual specific variance in addition to the common variance described by the domains (Costa & McCrae, 1995). The NEO-PI-R is currently the most widely used and the most researched marker of the FFM, and it has been demonstrated to outperform alternative instruments in comparative analyses (Ostendorf & Angleitner, 1994; for critical remarks, see Block, 1995). It is meant to cover the entire personality sphere, as described by the FFM.

Integrity Tests. Integrity was measured by the overt and personality-based part of the Inventar berufsbezogener Einstellungen und Selbsteinschätzungen (IBES, formerly labeled FES, Marcus, in press). The IBES was originally developed as a research instrument to allow for integrity test research in German-speaking countries but, after extensive validation, it is now in the process of publication by a leading German test publisher. It is primarily modeled after prototypical themes repeatedly identified in prominent U.S. integrity tests of both types (see Marcus, in press). The overt part of the IBES contains 60 items assigned to four subscales (general trust; perceived counter-productivity norms; rationalizations; behavioral intentions fantasies). The personality-based part has 55 items, divided into the component themes of manipulativeness,

trouble avoidance, positive self-concept, reliability/dependability, and stimulus seeking. All but two of these thematic scales are very similar to the content of typical U.S. integrity tests (e.g., Wanek *et al.*, 2003). The exceptions are general trust (a more general attitudinal scale that split off an earlier form of the counter-productivity norms scale), and manipulativeness, a trait found to be related to interpersonal forms of counter-productive behavior but typically not included in U.S. integrity tests (see Marcus, 2000, in press, for detailed discussions of the rationales involved in developing the IBES). All items are coded on a five-point Likert-type scale of endorsement.

In a series of studies summarized in the test manual (Marcus, in press), the IBES has been shown to yield acceptable levels of internal consistency (mean $\alpha = .91$ for the overt, .85 for the personality-based part, and .73 for the subscales) and retest reliability (mean r across scales = .81 with a 5-month time interval), an acceptable to good fit for one-factorial models at the item-level within single subscales (root mean square) error of approximation (RMSEAs; between .02 and .09), to show relationships with a wide variety of outside variables (e.g., cognitive ability, occupational interests, social desirability, and a wide range of personality constructs) that closely parallel findings with U.S. integrity tests, and to yield similar test-taker reactions as found with integrity tests in North America. Moreover, the two parts of the IBES consistently correlated with counter-productive work behaviors at comparable levels as reported in prior meta-analyses on integrity tests. Values range from $r = .31$ to .56 for the two IBES parts across eight samples (total $N > 1400$), including students, apprentices, employees, and employees instructed to act as applicants. In addition, the IBES correlated at about $r = .20$ with both academic grades and supervisory ratings of job performance in students, apprentices, and employees, respectively, and at about .30 with job satisfaction in various groups of employees. Whereas the criterion-related validity was almost unaffected by setting, the mean correlation between IBES subscales was $r = .22$ under standard instructions but .35 in simulated applicant settings, indicating considerable inflation in the latter case.

Further, the conceptual distinction between the two IBES parts is confirmed by the results of two recently conducted studies (details are available from the first author). In the first study based on three samples (German students, German employees, and Canadian students who took an English IBES translation), the personality-based part of the IBES was found to correlate at $r = .72$, .73, and .74, respectively, with a personality-based integrity test composed of CPI items (Hakstian *et al.*, 2002), while the overt part of the IBES showed slightly, but consistently lower correlations of .65, .67, and .57 in the same samples. The second study was based on two samples (German students, German employees) and used an overt integrity test as criterion (German translation of Ryan & Sackett, 1987; the same test as used by Lee, Ashton, & de Vries,

2005). The personality-based IBES part showed convergences of $r = .41$ and $.51$, respectively, while the overt IBES correlated at $.80$ and $.82$ with the American overt test.

Thus, there is considerable evidence from both construct and criterion-related validation studies with the IBES that consistently supports its equivalence with North-American integrity tests with respect to a multitude of findings from decade-long research in various settings.

Results

Study Descriptives

Means, standard deviations, and Cronbach's α reliabilities of all subscales included in the following analyses are reported in Table 2, along with the integrity-NEO correlations and the relationships between the former subscales (a full matrix of all 1035 coefficients is available upon request). It is evident that the correlational pattern of the personality-based IBES part (right part of Table 2) shows much more variation than that of the overt part, across integrity subscales as well as across NEO facets within the same domain. Accordingly, overt subscales display a relatively homogeneous pattern of positive associations among each other whereas the components of the personality-based test do not.

CFAs

Each scales' unidimensionality has been established by preceding analyses (see measures section). "Parallel" subscale halves were entered as parcels of observed variables into subsequent CFAs. These were formed by ordering items after the loadings of the respective subdimensions and then alternately assigning them to the scale halves. This procedure led to indicators in the form of a "partial disaggregation model" (Bagozzi & Heatherton,

1994). As opposed to single items, this reduces model complexity, approximates metric scale quality, takes advantage of the higher reliability and typically improved distributional properties of composites, and at the same time permits specification of a structural (latent higher-order level) model. Because of these desirable properties, parceling has been strongly recommended in a major review on structural equation modeling (MacCallum & Austin, 2000). None of the indicators exceeded values of 1 for skewness or kurtosis, which are critical for the robustness of maximum likelihood (ML) parameter estimation against deviations from multivariate normality (Muthén & Kaplan, 1985).

For each integrity test, a model with one higher-order factor, as assumed in the superfactor hypothesis, and a model with freely correlated subdimensions (measurement model), representing one portion of the several-facets hypothesis, were tested. The former specification is nested in the latter one, allowing for a comparative examination of model fit. Fit indices for these specifications are presented in Table 3.

There is a marked difference of the results between the two types of integrity tests. For personality-based subscales, model 2 fits reasonably well according to most indices reported in Table 3. When a higher-order factor was specified, however, we found either non-convergence or improper solutions (Heywood cases, i.e., negative variances), depending on the program used. According to Boomsma and Hoogland (2001), the causes of non-convergence and Heywood cases are almost identical, among them being insufficient sample size, number of indicators, model misspecification (poor fit), and low factor loadings. Evidence for the latter two explanations is revealed from inspection of the subscale intercorrelations in Table 2, where coefficients appear to vary unsystematically around 0. Increasing the number of indicators to solve the problem (Bollen, 1989) by replacing scale halves

Table 3. Fit statistics for confirmatory factor analyses of integrity test subdimensions

Model	Df	χ^2	GFI	AGFI	NFI	RMSEA	ECVI
Personality-based test							
1. Null model	45	677.56					3.29
2. No "integrity"-factor (correlated subscales)	25	39.30 ($p = .034$)	.97	.93	.94	.052	.47
3. Higher-order "integrity"-factor	Non-convergence or improper solutions						
Overt test							
4. Null model	28	841.23					4.04
5. No "integrity"-factor (correlated subscales)	14	51.06 ($p < .001$)	.95	.86	.94	.11	.45
6. Higher-order "integrity"-factor	16	65.46 ($p < .001$)	.93	.84	.92	.12	.50

Notes: $N = 213$; GFI, goodness-of-fit index; AGFI, adjusted goodness-of-fit index; NFI, normed fit index; RMSEA, root mean square error of approximation; ECVI, expected cross validation index.

by items led to a superfactor which loaded negatively ($-.51$) on one of its five subcomponents, very low ($-.04$, $.12$) on two others, and highly positive ($.52$, $.80$) on the remaining two facets. As this would describe a factor that could certainly not be interpreted as a meaningful construct, we had to reject the hypothesis of a general factor driving the relationships among personality-based IBES subscales. Without such a factor at this level, any interpretation of more abstract dimensions at higher levels of the hierarchy would be meaningless and is therefore not further examined.

For the overt test, by contrast, model 5 (correlated subscales) exhibited an only marginally acceptable fit to the data. But here, additional specification of a general factor, while leading to a highly significant increase in χ^2 ($\Delta\chi^2(2) = 14.4$; $p < .001$), only slightly impairs most other fit indices, indicating a significant yet not substantial decrease in fit. Moreover, the higher-order factor had substantial loadings on all subcomponents, ranging from .62 for behavioral intentions to .87 for rationalizations. This is again in accordance with subscale intercorrelations, which were homogeneously positive and substantial for the overt test (cf. Table 2). We therefore tentatively examined the full superfactor hypothesis for this type of integrity test by incorporating the FFM dimensions of emotional stability, agreeableness, and conscientiousness.

For this purpose, loadings of the general factor on its integrity subdimensions were fixed to the values found in the preceding isolated analysis. This should prevent a shift in the meaning of this factor because of additional parameters. Then, a model was specified where this overt integrity test dimension was allowed to load freely on the three relevant FFM domains. These three domains were not measured by their facets, as originally intended, but by domain halves constructed in the same manner as the integrity subscale halves. This was necessary because the hierarchical NEO model of domains and facets did not meet even the most liberal standards of model fit, as has been found in other CFA examinations of this inventory before (Borkenau & Ostendorf, 1990; Church & Burke, 1994; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996; Parker, Bagby, & Summerfeldt, 1993). Figure 2 provides a graphic representation of the exact specifications along with ML estimates of factor loadings (model fit: $\chi^2(82) = 211.20$, $p = .00$; RMSEA = .086; goodness-of-fit index = .88; adjusted goodness-of-fit index = .84; normed fit index = .89; CFA = .93).

It is again not easy to decide whether the data support or falsify the superfactor hypothesis with respect to overt tests. Model fit fell short of the acceptable range according to most indices, but in most cases only marginally so. It is usually hard to confirm models of some complexity – there were 82 degrees of freedom in this case – by means of CFA (relaxation of the fixed loadings on integrity subdimensions, however, did not improve model fit). Thus, the present data may best be interpreted as indicating marginal,
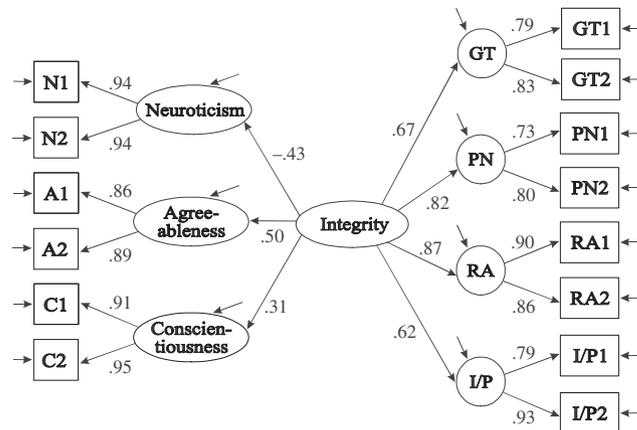
Figure 2. Graphic representation for the confirmatory factor analysis of the full *g*-factor model. Integrity = over test *g*-factor. For integrity subscale abbreviations, see Table 3. Coefficients show loadings (all loadings are significant at *p* < .01) for the completely standardized solution. Loadings on integrity subfactors are fixed to the values estimated in the isolated CFA, as reported in Table 3, model 6.

yet not conclusive, support for the existence of a higher-order factor comprised of overt integrity subdimensions and three of the big five. This cautious interpretation is corroborated by the loadings on the FFM factors, which are plausible in the light of earlier findings but not very substantial.

Multiple Regressions

A comparative test of the several-facets hypothesis and the superfactor hypothesis is approached by regressing both integrity test scores on selected NEO facets. If the latter hypothesis holds, aggregate integrity scale scores should be better predicted by a composite of all facets making up the relevant dimensions of ES, A, and C, whereas superior prediction by a more parsimonious set of facets would support the former position. A more conservative approach to test the superfactor hypothesis is provided by constraining beta weights for facets within each of the big five to equality. Table 4 provides the results for all multiple regressions, including both traditional measures of explained variance and SEM-specific indices of overall model fit. These types of statistics need to be treated separately. At the core of the competitive analyses are the overall fit indices expected cross validation index (ECVI) and CAIC. Both are non-redundant statistics designed to directly compare substantive models based on the same covariance matrix but not nested sequentially (as opposed to the χ^2 difference test), and both take parsimony (degrees of freedom) into account (cf. Jöreskog & Sörbom, 1993). An isolated examination of the several-facets hypothesis will emphasize R^2 and its variants as compared with the saturated model.

Table 4. Fit indices for multiple regression analyses by type of integrity test regressed on the relevant NEO-facets for both hypotheses tested

Model	df	χ^2	p for χ^2	CAIC	ECVI	RMSEA	GFI	AGFI	NFI	R^2	adj. R^2
Personality-based test as criterion											
Saturated model				3155.20	5.45					.81	.78
1.1a: Several-facets/free parameters	20	38.44	.0078	3066.42	5.44	.071	.99	.74	.99	.78	.76
1.2a: Superfactor/free parameters	12	65.99	.00	3144.85	5.68	.16	.98	.30	.98	.74	.72
1.1b: Several-facets/constrained parameters	26	59.07	.00022	3048.87	5.49	.084	.98	.71	.99	.75	
1.2b: Superfactor/constrained parameters	27	147.44	.00	3130.89	5.96	.16	.97	.42	.96	.62	
Overt test as criterion											
Saturated model	0			3155.20	5.45					.56	.49
2.1a: Several-facets/free parameters	23	28.57	.19	3037.46	5.35	.036	.99	.83	.99	.50	.49
2.2a: Superfactor/free parameters	12	18.02	.11	3096.89	5.42	.053	.99	.78	1.00	.52	.48
2.1b: Several-facets/constrained parameters	25	39.41	.034	3035.58	5.39	.056	.99	.79	.99	.47	
2.2b: Superfactor/constrained parameters	27	89.67	.0000001	3073.11	5.65	.11	.98	.60	.98	.33	

Notes: $N = 213$. ECVI and CAIC are comparative indices; best models according to these indices are italicized. All other indices are better interpreted with respect to the individual models they test. R^2 = squared multiple correlation, adj. R^2 = adjusted R^2. For exact model specifications, see Table 1.

To first focus on R^2, analyses revealed that the traits selected to test the several-facets hypothesis account for almost as much variance as does the full set of 30 NEO facets, particularly when adjusted for capitalization on chance. In addition, comparative fit indices indicate that restriction of parameters for irrelevant facets to 0 is justified according to the present data. Therefore, it might be concluded that a restricted set of narrow traits, taken from a broad range of higher-order dimensions, is sufficient to describe what integrity tests measure. However, this assertion has to be qualified with respect to the overt test for which about half of the variance remained unexplained by NEO facets (even when all 30 are entered). It would appear, then, that the overt integrity test used in this study is partially tapping into constructs that are not incorporated in Costa and McCrae's version of the FFM, and that only few FFM components are significant for its understanding.

By direct comparison of model fit, the several-facets hypothesis clearly outperformed the superfactor specifications according to any index for the personality-based test, particularly when additional equality constraints are introduced. The several-facets hypothesis also faired better than the superfactor hypothesis with the overt test, although, at least in the more liberal specification (model 2.2a), this difference was much less pronounced. Models 2.2a and 2.1a are also the only directly comparable pair of models for which the 90% confidence intervals of the ECVI overlap, indicating a non-significant difference in fit. This result, however, is qualified by the fact that the several-facet specification is much more parsimonious in this case (eight vs. 18 parameters). Thus, we found more support for the several-facets hypothesis with both types of integrity tests, but we could not reject the superfactor hypothesis with confidence for the overt integrity test. All in all, the difference by type of integrity test in multiple regressions closely parallels the results found with the CFA approach reported earlier.

Discussion

The present study was designed to clarify the somewhat inconclusive previous evidence on two competing conceptualizations of the construct(s) measured by integrity tests. In general, the results are quite unequivocal with respect to the distinction between types of integrity tests.

General Factor of Integrity

The first step of the analyses concerned the dimensionality of single integrity tests, a crucial yet neglected point in the original derivation of the superfactor hypothesis. The present data confirm a general integrity dimension to some degree for the overt test, but they clearly disconfirm a higher-order factor for the personality-based subscales. Thus, there appears to be a marked difference by test type

with respect to dimensionality. There is also evidence from the correlational analysis in Table 2 that a similar pattern of correlations for both types of integrity tests with FFM dimensions, which Wanek *et al*. (2003) found for a number of instruments, can be accompanied by notable differences at the facet level of the FFM. Thus, there appears to be good reason to treat overt and personality-based tests separately in the following parts of this discussion. Before doing so, we compare results of the present study with previous findings on the nature of integrity tests.

The lack of confirmation for an overall integrity factor directly contradicts the findings reported by Hogan and Brinkmeyer (1997), who found a good fit for a model with one higher-order factor based on component scores of overt and personality-based tests. There are several potential explanations for this apparent difference. First, Hogan and Brinkmeyer used substantially abbreviated versions of commercial integrity tests. This could have a profound effect on a test's dimensionality. For example, the original version of the personality-based integrity test used by Hogan and Brinkmeyer was about four times as long as the version they had actually administered, but the internal consistency of the abbreviated version was even higher ($\alpha = .75$ vs. .63; cf. Hogan & Hogan, 1989) – a finding strongly pointing to a much more homogeneous structure of the short form. The present authors used considerably longer instruments, but these were developed in a different culture. Despite the consistent evidence of equivalence of the IBES with U.S. integrity tests collected so far, direct comparisons involving more than one instrument of each type would be clearly desirable.

Further, Hogan and Brinkmeyer's sample was more than 10 times as large as that of the present study. Thus, any possible artifacts because of sampling error are more likely to be a problem for the present investigation. Still another important difference is the fact that the former sample consisted of actual job candidates, whereas students served as participants for the present research. Our choice of a student sample was based on previous findings that personality scale intercorrelations tend to be severely inflated in applicant samples. If this is the case (see introduction and measures section for evidence, but see also Smith & Ellingson, 2002), it is apparent that confirmation of a general factor becomes more likely in applicant settings. Our approach to employ an anonymous laboratory setting is in accordance with traditional research on the structure of personality, a tradition that revealed – among many other things – the current image of the FFM. However, the psychological meaning of test scores obtained under the motivational conditions of an applicant setting is certainly of interest in its own right.

Personality-Based Integrity Test

For the present purpose of investigating the internal structure of integrity tests in an anonymous setting, it has to be concluded that the personality-based instrument used here does not measure a homogeneous construct. In actual selection situations, one may confirm a higher-order factor using such instruments, but this factor appears to be much like the general factor detected in almost any kind of evaluative personality test. By logical reasoning, then, one is inclined to seek an explanation for the incremental validity of personality-based integrity tests over mainstream personality inventories in those aspects, which are unique to each.

As confirmed by the multiple regression analyses, the several-facets hypothesis provides one plausible basis for deducing such an explanation. Personality-based integrity tests sample selectively from the universe of traits, whereas mainstream inventories usually attempt to cover the entire personality sphere. By the former strategy, it is possible to choose only elements that are relevant for the criterion to be predicted and omit any irrelevant portions of variance. While this, of course, is not a novel way to develop valid tests,[3] its results can tell us much about the nature of the criterion, given that prediction is most successful when the predictor is modeled after the criterion (e.g., Hogan & Roberts, 1996; Schmidt & Kaplan, 1971). The image that the several-facets view of integrity tests would present of counter-productivity, the main target criterion of these tests, is that of a complex criterion. Complex, however, is not synonymous with "broad": while many diverse traits may lead to propensities for misconduct at work, these, it seems, are to be sought at the level below rather than above the big five dimensions (e.g., Ashton, 1998). Some of these lower-order traits were selected for this study on the basis of rational judgment, but clearly more research is needed to corroborate or perhaps alter the present selection and provide a more precise weighting. Completion of this task may profit from theories and research on the links between personality and criminal or deviant conduct (e.g., Marcus & Schuler, 2004), and would require the assessment of appropriate criteria.

Overt Integrity Test

The image of complexity changes considerably when one tries to derive it from the findings on overt integrity tests, developed to predict the same criteria but usually in terms of more narrow behavioral categories like theft or substance abuse. This suggests that the picture just described may not be complete. The existence of a general integrity factor, as implicitly assumed in the superfactor hypothesis, received at least marginal support from the current CFA on overt subscales. Moreover, the same factor showed loadings in the range of .30 to .50 and a marginally acceptable fit when the FFM dimensions of emotional stability, agreeableness, and conscientiousness are incorporated into the model (including extraversion and openness decreased model fit substantially while loadings for these factors are close to 0). These findings

may be interpreted as partial confirmation of the super-factor hypothesis for this type of integrity tests. However, some details of our data analyses seem to qualify this conclusion.

First, support for the superfactor hypothesis was not nearly as unequivocal as it was for the several-facets view with the personality-based test, particularly when examined in direct comparison by multiple regressions. Here, the more parsimonious several-facets model faired slightly better and this gap enlarged substantially when equality constraints were introduced. The competitive models served only as a proxy in the case of the superfactor hypothesis, but one would expect a better fit and, in particular, a larger proportion of explained variance if N, A, and C, or even their higher-order communalities, were comprehensive explanations for what integrity tests measure.

The impression of an incomplete explanation was further corroborated by the only moderate magnitudes of loadings for the overt integrity factor on N, A, and C, as found in the CFA. In the multiple regression analyses, even the full variance of all 30 facets was unable to account for more than about half of the variance in the overt integrity test in a multiple regression. It would appear, then, that neither a superfactor of FFM dimensions nor a combination of their facets comprehensively describes what the overt part of the IBES measures. Both positions had their merits with this instrument but there is obviously something missing.

The present results point to the conclusion that this missing piece is to be sought outside the FFM. For this purpose, it is useful to distinguish two often-confused features of this model of personality, its robustness (or generalizability), and its comprehensiveness. The generalizability of the FFM across cultures, populations, questionnaires, methods, etc. has been demonstrated in countless studies (see, e.g., Digman, 1990; John, 1990, for reviews). Thus, to our knowledge, it has a more general validity than any alternative model. Its comprehensiveness, however, has been questioned from different perspectives (e.g., Andresen, 1995; Block, 1995; Hough, 1992). The assertion that there are no substantive dispositions beyond the big five relies heavily on the fundamental assumption of the lexical approach that all descriptions of human character are manifested in natural language (e.g., Saucier & Goldberg, 1996). Notably, the most prominent advocates of this approach have recently proposed a number of additional factors, derived by extending their analysis to adjectives which are not "trait-like" in a narrow sense (Saucier & Goldberg, 1998). Some of the dimensions that fall outside the FFM are attitudinal constructs, according to these authors (Paunonen & Jackson, 2000, reanalyzed the same data and identified several further adjective clusters beyond the big five, including "honesty").

While this line of research has also led to alternative explanations of what integrity tests measure that are closer to the FFM (e.g., Lee *et al.*, 2005, suggest that an additional honesty factor at the same level as the big five is at the core of the construct measured by overt integrity tests[4]), the general distinction of traits vs. attitudes appears promising as an aid to understanding what distinguishes the two types of integrity tests. These concepts have their historical roots in the distinct traditions of personality assessment and social psychology, respectively, but there is no doubt that both have much in common (temporal stability; dispositional constructs meant to explain human behavior). It is also evident that both are not simply synonyms (for instance, attitudes describe a relationship between a person and an object, whereas traits mainly describe the self; for discussions of this topic, see Ajzen, 1988; Sherman & Fazio, 1983) and, hence, may best be conceptualized as related but not completely overlapping. This is precisely the image of relationships between both types of integrity tests that may be drawn from the results presented herein and in previous research (e.g., Ones, 1993).

In fact, it has long been suggested that the now commonplace distinction of overt vs. personality-based be replaced by the more theoretically sound dichotomy of attitudinal vs. trait-based (Marcus *et al.*, 1997). The latter might be taken as synonymous to personality-based, if that means "based on traditional personality inventories," which are themselves based on the trait approach to personality. The former term seeks to replace overtness, a feature of single items with the connotation of fakability, by the theoretical concept of attitudes, which refer to the constructs measured on a similar conceptual level as traits.

There are several reasons why this distinction might be regarded as fruitful for our theoretical understanding of integrity tests. First, it appears to provide a more balanced dichotomy and, thus, one that is less susceptible to confusing theoretical with psychometric issues. Second, the "overtness" of single items appears to vary substantially within overt tests (Dwight & Alliger, 1997) and may therefore not adequately describe an entire scale or even test category. Third, the most explicit reference to theoretical foundations of these instruments provided by publishers of overt tests was to Ajzen and Fishbein's (1977) congruence theory of attitude–behavior relationships (Jones, 1991; Jones & Terris, 1991). Fourth, the results of the present study indicate that there is a marked difference in the degree to which both kinds of tests fit in the currently most researched framework of personality traits, the FFM.

The present findings represent another piece of evidence that personnel psychology has something to discover not only beyond cognitive abilities but also beyond the FFM. Future research should go further than aggregating or disaggregating big five dimensions at different levels of the hierarchy. This may lead to concepts that are essentially different from the traits currently organized within the FFM. Attitudes may be one candidate for such an extension.

Study Limitations

The conclusions drawn on the present data are, of course, not only limited by the scope of the investigation but also by a number of additional shortcomings.

Recent simulation studies by Jackson (2003) support the notion that the "observations per parameter"-ratio is a meaningful criterion to assess the adequacy of sample size. Kline (2004) corrects older advice (e.g., Bentler, 1985, recommended a ratio of 5:1) by suggesting a 10:1 ratio as a rule of thumb. The most complex models (operationalized by number of estimated parameters) are model 2 (30 parameters) and 5 (22 parameters) in Table 3. Given our sample size of $N = 213$, the ratios for these models are 7.1:1 and 9.7:1, respectively. Thus our sample was not sufficiently large to meet recent standards for a satisfactory "observation per parameter"-ratio with our most complex models, though perhaps only marginally so.

Other potential limitations refer to the generalizability of our results. First, unknown cultural differences may have affected our findings. However, recently conducted studies (briefly summarized in the method section) strongly support the generalizability of findings on the constructs measured by the IBES at least across two countries (Germany vs. Canada). Moreover, these and other findings reported in the manual (Marcus, in press) are highly similar to results obtained with established integrity tests in the U.S. Still, the paucity of research on integrity tests outside the English speaking world is one of the major gaps to be filled in the future.

It has also been mentioned that conducting the study on students instead of applicants cannot automatically be seen as a drawback when the focus of the study is on theoretical issues. One useful compromise between the lack of representativeness of university students and the potentially distorting motivational state of actual job candidates may be to recruit a sample of current employees who are guaranteed anonymity. Unfortunately, this was not possible for the present study, but the studies mentioned in the preceding paragraph also demonstrated that notable differences between student and employee samples are unlikely to occur as long as the test situation is similar. Nevertheless, future research should also address whether the testing situation affects the constructs measured by integrity tests in similar ways as it affects FFM measures.

Further, our findings may have been affected by specifics of the instruments used. First, we used only one model for FFM dimensions and facets, the NEO model. Research employing different operationalizations (e.g., Caprara, Barbaranelli, Borgogni, & Perugini, 1993; Hogan, 1986) may lead to different results but the NEO approach appeared to be currently the most elaborated one at the facet level. Secondly, we also employed only one integrity test per type, and the two tests were parts of a single inventory. The observed correlation between the two measures was $r = .47$, which is similar to values reported by Ones (1993) for the most widespread U.S. integrity tests of both types. Moreover, if treating parts of one inventory as separate measures had affected results in the present study, it would have obscured differences between the two parts rather than the opposite. Furthermore, beyond the direct evidence of prototypicality of the IBES parts for their domains cited in the method section, we could compare the pattern of IBES–NEO correlations with those reported in earlier studies on personality-based U.S. integrity tests (Costa & McCrae, 1995; Hakstian et al., 2002). If we correlated the vectors of NEO-facets/integrity correlations in our study with those in the U.S., the observed vector correlations are: $r = .76$ between IBES personality-based (pb) and Hakstian et al.'s male sample (.67 with IBES overt); .97 between IBES pb and Hakstian et al.'s female sample (.89 with IBES overt), and .89 between IBES pb and Costa and McCrae (.83 with IBES overt). The respective values among the various U.S. samples range from .83 to .91. This seems to indicate a high degree of similarity between the present and previous U.S. findings on that issue.

One final shortcoming of the present study, as far as they are mentioned here, was dictated by the conceptual incompatibility of the two hypotheses tested. None of the direct comparisons represented both positions completely adequate at the same time. Given the methodological differences of our two approaches, however, it is note-worthy that both led to essentially the same conclusions regardless of test type investigated.

Conclusion

In the present article, two hypotheses on the relationships between integrity tests and the FFM were reviewed and empirically examined. One, labeled the superfactor hypothesis, states that integrity tests of any kind measure a general factor of personality, comprised of three of the big five dimensions. The second view, called the several-facets hypothesis, asserts that integrity tests are composed of several narrow personality traits, which need not be related among each other via any higher-order factor. Results of this study, the first one to test both positions in one data set, provided substantial support for the several-facets hypothesis with respect to a personality-based integrity test. By contrast, the superfactor hypothesis had to be rejected for this type of integrity test. With overt instruments, both positions received limited support but none was able to account for a comprehensive understanding of its scale scores. Findings point to the conclusion that a clearer conceptual distinction between both types of integrity tests may be a fruitful starting point for a more comprehensive understanding of the construct of integrity. Future studies on this issue may employ different operationalizations of the FFM at the facet level, different integrity measures,

different populations in different contexts, and incorporate the assessment of job-relevant criteria.

Acknowledgements

Notes

1. In more recent writings, Ones and colleagues have expressed the view that they regard the construct of integrity as multifaceted (e.g., Ones & Viswesvaran, 2001). Thus, it is important to note that the present review represents our interpretation of early formulations of and research on the superfactor hypothesis, not what the original authors actually think.
2. Facets, although theoretically not ideal for testing the superfactor hypothesis, are used to make analyses directly comparable. Results for multiple regressions with three domains as predictors are available from the first author. They do not change the conclusions based on the results presented herein.
3. We would not equate the described strategy with the often-criticized "blind" empirical criterion-keying method. It is well possible to base test construction on substantive theoretical reasoning but still select several heterogeneous constructs solely on the grounds that they all are related to the criterion of interest, but for different reasons.
4. We believe that Lee *et al.*'s (2005) suggestion is different but not contradictory to the distinction between traits and attitudes stressed in the present paper. Like us, Lee and colleagues emphasized the difference between the two types of integrity tests, and the insufficiency of the FFM to account for the construct measured by overt tests. This is certainly an interesting avenue for future research.

References

Ajzen, I. (1988) *Attitudes, personality and behavior*. Stony Stratford, UK: Open University Press.

Ajzen, I. and Fishbein, M. (1977) Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, **84**, 888–918.

Andresen, B. (1995) Risikobereitschaft (R) – der sechste Basisfaktor der Persönlichkeit: Konvergenz multivariater Studien und Konstruktexplikation. ["Risk preference" (R) – the sixth basic factor of personality: Convergence of multivariate studies and construct explication]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, **16**, 210–236.

Ashton, M.C. (1998) Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior*, **19**, 289–303.

Bagozzi, R.P. and Heatherton, T.F. (1994) A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling*, **1**, 35–67.

Barrick, M.R., Mount, M.K. and Judge, T.A. (2001) Personality and performance at the beginning of the new millenium: What do we know and where do we go next? *International Journal of Selection and Assessment*, **9**, 9–30.

Becker, T.E. (1998) Integrity in organizations: Beyond honesty and conscientiousness. *Academy of Management Review*, **23**, 154–161.

Bentler, P.M. (1985) *Theory and implementation of EQS: A structural equations program*. Los Angeles: BMDP Statistical Software.

Block, J. (1995) A contrarian view on the five-factor approach to personality description. *Psychological Bulletin*, **117**, 187–215.

Bollen, K.A. (1989) *Structural equations with latent variables*. New York: Wiley.

Bollen, K.A and Lennox, R. (1991) Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, **110**, 305–314.

Boomsma, A. and Hoogland, J.J. (2001) The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit and D. Sörbom (Eds), *Structural equation models: Present and future* (pp. 139–168). Chicago: Scientific Software International.

Borkenau, P. and Ostendorf, F. (1990) Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences*, **11**, 515–524.

Borofsky, G.L. (1992) Assessing the likelihood of reliable workplace behavior: Further contributions to the validation of the Employee Reliability Inventory. *Psychological Reports*, **70**, 563–592.

Borofsky, G.L. (1993) *User's manual for the employee reliability inventory screening system*. Libertyville, IL: Wonderlic.

Brooks, P. and Arnold, D.W. (1989) *Reid Report Examiner's Manual*. Chicago, IL: Reid Psychological Systems.

Brown, R. and Barrett, P. (1999, June) Differences between applicant and non-applicant personality questionnaire data. Paper presented at the 1999 British Psychological Society's Test User Conference, Scarborough, UK.

Caprara, G.V., Barbaranelli, C., Borgogni, L. and Perugini, M. (1993) The big five questionnaire: A new questionnaire for the measurement of the five factor model. *Personality and Individual Differences*, **15**, 281–288.

Church, A.T. and Burke, P.J. (1994) Exploratory and confirmatory tests of the big five and Tellegen's three- and four dimensional models. *Journal of Personality and Social Psychology*, **66**, 93–114.

Collins, J.M. and Bagozzi, R.P. (1999) Testing the equivalence of the socialization factor structure for criminals and noncriminals. *Journal of Personality Assessment*, **72**, 68–73.

Collins, J.M. and Gleaves, D.H. (1998) Race, job applicants, and the five-factor model of personality: Implications for black psychology, industrial/organizational psychology, and the five-factor theory. *Journal of Applied Psychology*, **83**, 531–544.

Costa, P.T. and McCrae, R.R. (1992) *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P.T. and McCrae, R.R. (1995) Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, **64**, 21–50.

Cunningham, M.R. and Ash, P. (1988) The structure of honesty: Factor analysis of the Reid report. *Journal of Business and Psychology*, **3**, 54–66.

Digman, J.M. (1990) Personality structure: emergence of the Five-Factor Model. *Annual Review of Psychology*, **41**, 417–440.

Digman, J.M. (1997) Higher-order factors of the big five. *Journal of Personality and Social Psychology*, **73**, 1246–1256.

Dwight, S.A. and Alliger, G.M. (1997) Reactions to overt integrity test items. *Educational and Psychological Measurement*, **57**, 937–948.

Ghiselli, E.E., Campbell, J.P. and Zedeck, S. (1981) *Measurement theory for the behavioral sciences*. San Francisco, CA: Freeman.

Gough, H.G. (1960) Theory and measurement of socialization. *Journal of Consulting Psychology*, **24**, 23–30.

Gough, H.G. (1971) The assessment of wayward impulse by means of the personnel reaction blank. *Personnel Psychology*, **24**, 669–677.

Gough, H.G. and Peterson, D.R. (1952) The identification and measurement of predispositional factors in crime and delinquency. *Journal of Consulting Psychology*, **16**, 207–212.

Hakstian, A.R., Farrell, S. and Tweed, R.G. (2002) The assessment of counterproductive tendencies by means of the California psychological inventory. *International Journal of Selection and Assessment*, **10**, 58–86.

Harris, W.G. (1987) A components analysis of a preemployment integrity measure: A replicated study. *Psychological Reports*, **60**, 1051–1055.

Harris, M.M. and Sackett, P.R. (1987) A factor analysis and item response theory analysis of an employee honesty test. *Journal of Business and Psychology*, **2**, 122–135.

Hogan, J. and Brinkmeyer, K. (1997) Bridging the gap between overt and personality-based integrity tests. *Personnel Psychology*, **50**, 587–599.

Hogan, J. and Hogan, R.T. (1989) How to measure employee reliability. *Journal of Applied Psychology*, **74**, 273–279.

Hogan, J. and Roberts, B.W. (1996) Issues and non-issues in the fidelity–bandwidth trade-off. *Journal of Organizational Behavior*, **17**, 627–637.

Hogan, R.T. (1986) *Hogan personality inventory manual*. Minneapolis, MN: National Computer Systems.

Hogan, R.T. and Jones, W.H. (1983) A role-theoretical model of criminal conduct. In W.S. Laufer and J.M. Day (Eds), *Personality theory, moral development, and criminal behavior* (pp. 3–21). Lexington, MA: Lexington Books.

Hough, L.M. (1992) The "big five" personality variables – construct confusion: Description versus prediction. *Human Performance*, **5**, 139–155.

Hough, L.M. and Schneider, R.J. (1996) Personality traits, taxonomies, and applications in organizations. In K.R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 31–88). San Francisco: Jossey-Bass.

Jackson, D.L. (2003) Revisiting sample size and number of parameter estimates: Some support for the N.q hypothesis. *Structural Equation Modeling*, **10**, 128–141.

John, O.P. (1990) The "big five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L.A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: Academic Press.

Jones, J.W. (1991) Attitude–behavior relations: A theoretical and empirical analysis of preemployment integrity tests. In J.W. Jones (Ed.), *Preemployment honesty testing: Current research and future directions* (pp. 89–98). Westport, CT: Quorum Books.

Jones, J.W. and Terris, W. (1991) Integrity testing for personnel selection: An overview. *Forensic Reports*, **4**, 117–148.

Jöreskog, K.G. and Sörbom, D. (1993) *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International.

Kline, R.B. (2004) *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kochkin, S. (1987) Personality correlates of a measure of honesty. *Journal of Business and Psychology*, **1**, 236–247.

Law, K.S, Wong, C.S. and Mobley, W.H. (1998) Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, **23**, 741–755.

Lee, K., Ashton, M.C. and de Vries, R.E. (2005) Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure. *Human Performance*, **18**, 179–197.

MacCallum, R.C. and Austin, J.T. (2000) Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, **51**, 201–226.

Marcus, B. (2000) *Kontraproduktives Verhalten im Betrieb: Eine individuumsbezogene Perspektive* [*Counterproductive behavior in organizations: An interindividual perspective*]. Göttingen, Germany: Verlag für Angewandte Psychologie.

Marcus, B. (in press) *Inventar berufsbezogener Einstellungen und Selbsteinschätzungen* [*Job-Related attitudes and self-evaluations inventory*]. Göttingen, Germany: Hogrefe.

Marcus, B. and Schuler, H. (2004) Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology*, **89**, 647–660.

Marcus, B., Funke, U. and Schuler, H. (1997) Integrity Tests als spezielle Gruppe eignungsdiagnostischer Verfahren: Literaturüberblick und metaanalytische Befunde zur Konstruktvalidität [Integrity tests as a specific group of instruments in personnel selection. A literature review and meta-analytic findings on construct validity]. *Zeitschrift für Arbeits – und Organisationspsychologie*, **41**, 2–15.

McCrae, R.R., Costa, P.T. and Piedmont, R.L. (1993) Folk concepts, natural language and psychological constructs: The California psychological inventory and the five-factor model. *Journal of Personality*, **61**, 1–26.

McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H. and Paunonen, S.V. (1996) Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology*, **70**, 552–566.

McFall, L. (1987) Integrity. *Ethics*, **98**, 5–20.

Murphy, K.R. and Lee, S.L. (1994) Does conscientiousness explain the relationship between integrity and job performance? *International Journal of Selection and Assessment*, **2**, 226–233.

Muthén, B. and Kaplan, D. (1985) A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, **38**, 171–189.

Ones, D.S. (1993) The construct validity of integrity tests. Unpublished doctoral dissertation. University of Iowa, Iowa City, IO.

Ones, D.S. and Viswesvaran, C. (1996) Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, **17**, 609–626.

Ones, D.S. and Viswesvaran, C. (2001) Integrity test and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, **9**, 31–39.

Ones, D.S., Schmidt, F.L. and Viswesvaran, C. (1994a) Do broader personality variables predict job performance with higher validity? Paper presented at the 1994 Conference of the Society for Industrial and Organizational Psychology, Nashville, TN, U.S.A.

Ones, D.S., Schmidt, F.L. and Viswesvaran, C. (1994b, July) Correlates of preemployment integrity tests. Paper presented at the 23rd International Congress of Applied Psychology, Madrid, Spain.

Ones, D.S., Viswesvaran, C. and Schmidt, F.L. (1993) 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.

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

Ostendorf, F. and Angleitner, A. (2003) *NEO-Persönlichkeitsinventar nach Costa und McCrae, revidierte Form (NEO-PI-R)*. Göttingen, Germany: Hogrefe.

Paajanen, G.E. (1988). The prediction of counterproductive behavior by individual and group variables. Unpublished doctoral dissertation. University of Minnesota, Minneapolis, MN.

Parker, J.D.A., Bagby, R.M. and Summerfeldt, L.J. (1993) Confirmatory factor analysis of the revised NEO personality inventory. *Personality and Individual Differences*, **15**, 463–466.

Paunonen, S.V. and Jackson, D.N. (2000) What is beyond the big five? Plenty!. *Journal of Personality*, **68**, 821–835.

Rosen, A.S. (1977) On the dimensionality of the California psychological inventory socialization scale. *Journal of Consulting and Clinical Psychology*, **45**, 583–591.

Ryan, A.M. and Sackett, P.R. (1987) Pre-employment honesty testing: Fakability, reactions of test takers, and company image. *Journal of Business and Psychology*, **2**, 248–256.

Sackett, P.R. and 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**, 787–829.

Sackett, P.R., Burris, L.R. and Callahan, C. (1989) Integrity testing for personnel selection: An update. *Personnel Psychology*, **42**, 491–529.

Saucier, G. and Goldberg, L.R. (1996) The language of personality: Lexical perspectives on the five-factor model. In J.S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 21–50). New York: Guilford.

Saucier, G. and Goldberg, L.R. (1998) What is beyond the big five? *Journal of Personality*, **66**, 495–524.

Schmidt, F.L. and Hunter, J.E. (1998) The validity and utility of personnel selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, **124**, 262–274.

Schmidt, F.L. and Kaplan, L.B. (1971) Composite vs. multiple criteria: A review and resolution of the controversy. *Personnel Psychology*, **24**, 419–434.

Schmit, M.J. and Ryan, A.M. (1993) The big five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, **78**, 966–974.

Schneider, R.J. and Hough, L.M. (1995) Personality and industrial/organizational psychology. In C.L. Cooper and I.T. Robertson (Eds), *International review of industrial and organizational psychology*, Vol. 10 (pp. 75–129). Chichester, UK: Wiley.

Schneider, R.J, Hough, L.M. and Dunnette, M.D. (1996) Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior*, **17**, 639–655.

Sherman, S.J. and Fazio, R.H. (1983) Parallels between attitudes and traits as predictors of behavior. *Journal of Personality*, **51**, 308–345.

Smith, D.B. and Ellingson, J.E. (2002) Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, **87**, 211–219.

Ullman, J.B. and Newcomb, M.D. (1998) Eager, reluctant, and nonresponders to a mailed longitudinal survey: Attitudinal and substance use characteristics differentiate responses. *Journal of Applied Social Psychology*, **28**, 357–375.

Wanek, J.E. (1995). The construct of integrity: Item level factor analysis of the dimensions underlying honesty testing and big-five measures of personality. Unpublished doctoral dissertation. University of Minnesota, Minneapolis, MN.

Wanek, J.E., Sackett, P.R. and Ones, D.S. (2003) Towards and understanding of integrity test similarities and differences: An item-level analysis of seven tests. *Personnel Psychology*, **56**, 873–894.

Wiggins, J.S. (Ed.). (1996) *The five-factor model of personality: Theoretical perspectives*. New York: Guilford Press.

Wilson, J.Q. and Herrnstein, R.J. (1985) *Crime and human nature*. New York: Simon & Schuster.