PERSONNEL PSYCHOLOGY
2007, 60, 1–34

PERSONALITY DIMENSIONS EXPLAINING RELATIONSHIPS BETWEEN INTEGRITY TESTS AND COUNTERPRODUCTIVE BEHAVIOR: BIG FIVE, OR ONE IN ADDITION?

BERND MARCUS
University of Western Ontario

KIBEOM LEE
University of Calgary

MICHAEL C. ASHTON
Brock University

Although the criterion-related validity of integrity tests is well established, there has not been enough research examining which personality constructs contribute to their criterion-related validity. Moreover, evidence of how well findings on integrity tests in North America generalize to non-English speaking countries is virtually absent. This research addressed these issues with data obtained from employees and students in Canada and Germany (total $N = 853$). Specifically, we tested the hypotheses that (a) Honesty–Humility, as specified in the HEXACO model of personality, is relatively more important than the Big 5 dimensions of personality in accounting for the criterion-related validity of overt integrity tests, whereas (b) the Big 5 are relatively more important in explaining the validity of personality-based integrity tests. These predictions were tested using 2 criteria (counterproductive work behavior and counterproductive academic behavior) as well as 2 overt and 2 personality-based integrity tests. We found evidence of the expected differences between types of integrity tests largely regardless of culture of the sample, specific test, criterion, or population under research, pointing to some degree of generalizability of findings in integrity testing research. Implications include theoretical refinements in research on integrity testing and encouragement of practical applications beyond North America.

An earlier version of this paper was presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX. This project was supported by Social Sciences and Humanities Research Council of Canada grants 410-2003-0946 and 410-2003-1835. We thank Birgit Schyns for translation of the HEXACO Personality Inventory into the German language and Diane Bielicki for translation of the IBES into the English language. We also thank Maren Bracht and Martin Kofler for their aid in translations of various measures into German and data collection in West Germany, and Michela Schröder for her help in collecting East German data.

Correspondence and requests for reprints should be addressed to Bernd Marcus, Department of Psychology & Management and Organizational Studies Program, University of Western Ontario, London, Ontario, N6A 5C2, Canada; bmarcus3@uwo.ca.

The validity of integrity tests in predicting various job-related criteria is well established through extensive research, as summarized more than a decade ago in one of the largest meta-analyses on instruments used in personnel selection (Ones, Viswesvaran, & Schmidt, 1993). Less well established, however, are the theoretical sources of integrity tests' criterion-related validities. Ones and Viswesvaran (2001) describe integrity tests as the prototype of what they labeled Criterion-Focused Occupational Personality Scales (COPS). On one hand, COPS are personality measures in that they tap into individual differences beyond the domain of cognitive abilities by means of standardized psychometric measurement. However, unlike standard measures of adult normal personality designed to describe individuals in terms of well-defined traits and structural models of personality, integrity tests and other COPS are mainly constructed to predict important work-related criteria. That is, the focus in test construction is on predicting specific target criteria rather than on measuring theoretically founded personality constructs.

Although the focus on criteria may enhance predictor–criterion relationships, it does not facilitate the advancement of theories of job behaviors unless the reasons for criterion-related validity are understood (e.g., Marcus, 2000; Ones, 1993). The label of "integrity," as well as a number of alternative labels (cf. Sackett & Wanek, 1996), was attached to a category of tests long after many of these tests were already in use. To understand the theoretical bases of actual integrity tests' validity, it would therefore be misleading to infer the meaning of integrity test scores deductively from any theoretical definition of integrity (e.g., Becker, 1998; McFall, 1987), as there is no actual link between those construct definitions and the tests that now share the construct's name. Instead, we take an inductive approach and try to infer the constructs measured by integrity tests from actual scores on these tests. In other words, we were not interested in the meaning of "integrity" as a construct; instead, we were interested in understanding the meaning of integrity test scores. To avoid further confusion, we refer to integrity *tests* or *test scores* as our target of interest in the remainder of this paper. Our major objective is to shed some new light on the bases, within descriptive models of personality, of integrity test scores' behavioral correlates. We focus on counterproductive behavior in the present paper because this is the core target criterion integrity tests are designed to predict. Further, recent evidence suggests that it is one of the major components of work behavior and performance (Rotundo & Sackett, 2002).

Despite the prominent role attributed to integrity tests in a number of fundamental issues surrounding the use of personality measures for personnel selection (e.g., Ones & Viswesvaran, 1996, 1998; Schmidt & Hunter, 1998; Schneider, Hough, & Dunnette, 1996), little research has

directly addressed the theoretical link between the constructs measured by integrity tests and the target behavior. There is a considerable volume of indirect evidence, though. Similar to other COPS (e.g., customer service or stress tolerance scales; cf. Ones & Viswesvaran, 2001), integrity test scores were found to correlate consistently, though at moderate levels, with three of the dimensions of the five-factor model of personality (FFM), namely Conscientiousness (C), Agreeableness (A), and Emotional Stability (ES) (Marcus, Funke, & Schuler, 1997; Ones, 1993). More recently, Wanek, Sackett, and Ones (2003) identified four broad themes common among the most widely used integrity tests and confirmed that C, A, and ES were those integrity test themes' major correlates within the FFM (but Wanek et al. also report moderate relations with Extraversion and Openness to Experience). As C, A, and ES are the very FFM dimensions that were also found to be associated with counterproductive behavior (Berry, Ones, & Sackett, in press; Salgado, 2002), it is tempting to seek the major source of integrity test score—counterproductivity relations in those traits.

Moreover, there is evidence from multiple studies that a higher-order factor, even more abstract than the FFM dimensions, underlies relationships between C, A, and ES (Digman, 1997). Digman interpreted this higher-order factor as a broad measure of the general success of socialization, indicating superego and learned internal restraint of impulses. Ones and Viswesvaran (2001) suggested that it could "be the most important trait that needs to be systematically measured among job applicants" (p. 37), but they carefully added a question mark.

The evidence in favor of C, A, and ES (or a common element underlying these dimensions) as the theoretical link between integrity test scores and counterproductive behavior appears striking indeed. However, there is also evidence that may give rise to a more skeptical position. For example, although the most robust correlate of integrity test scores among the five factors is C (Ones, 1993), Murphy and Lee (1994a) found little support for the hypothesis that C explains the validity of integrity tests in predicting job performance in a meta-analytic study. In addition, a number of primary studies found that integrity tests correlated substantially differently with specific facets of FFM dimensions within the same factors (Costa & McCrae, 1995; Hakstian, Farrell, & Tweed, 2002; Marcus, Höft, & Riediger, 2006; Murphy & Lee, 1994b), suggesting that the construct(s) measured by integrity tests may be sought at the level *below* the Big Five, rather than above the dimensions. Furthermore, relationships between counterproductive behavior and FFM dimensions are low to moderate at best. Meta-analytically corrected correlations of ES, A, and C with general counterproductive behavior ranged from $\rho = -.06$ to $-.26$ in Salgado's (2002) study. More recently, Berry et al. (in press) reported values up to $\rho = -.46$ between A and interpersonal deviance, and $-.42$

between C and organizational deviance (cf. Robinson & Bennett, 1995), respectively; though most correlations were substantially lower, integrity test score–FFM relations are not much higher (Marcus et al., 1997; Ones, 1993). This casts some doubt on the power of C, A, and ES for explaining the substantial integrity–criterion correlations.

The most recent challenge of the assumption that mainly C, A, and ES underlie integrity test score–criterion relationships comes from applied work on the HEXACO model of personality (Lee, Ashton, & de Vries, 2005; Lee, Ashton, & Shin, 2005). Whereas the vast majority of the previous research on personality foundations of integrity tests focused on variants of the Big Five and the FFM (though not always on the dimensional level), the HEXACO model points to an explanation not covered within the five-factorial space. HEXACO is an acronym for six, not just five, factors of personality, whereby the latter five dimensions relatively closely resemble the content of the Big Five (Emotionality or low ES, eXtraversion, Agreeableness, Conscientiousness, Openness to experience), whereas the first factor, Honesty–Humility (H–H), adds an essentially new component. This factor is marked by adjectives like *honest*, *fair*, *sincere*, or *loyal* versus *greedy*, *conceited*, *pretentious*, and *sly*. H–H is typically modestly related to the FFM dimensions of A and C (cf. Table 1) and can be empirically and conceptually clearly distinguished from those dimensions (Ashton et al., 2004). As opposed to task-related traits like organization or self-discipline defining FFM-C, H–H primarily refers to moral conscience. As compared with FFM-A, (low) H–H is more closely related to traits like primary psychopathy or Machiavellianism, which describe tendencies toward socially problematic, egotistic behaviors and attitudes (cf. Lee, Ashton, & de Vries, 2005; Lee, Ashton, & Shin, 2005), and does not generally include content related to hostility or toughness.

The HEXACO model differs from other models proposing a personality space larger than that of the Big Five/FFM in some important ways (see also Ashton & Lee, in press, for empirical and theoretical foundations of the HEXACO model). First, it was derived with the same approach as was the original Big Five framework, using factoring comprehensive sets of trait-descriptive adjectives found in natural languages. Studies with a wide variety of cultures converged on the conclusion that six semantically consistent factors emerged from sets of trait adjectives in natural languages—including English and German, which are spoken by participants in the present research (Ashton, Lee & Goldberg, 2004; Ashton et al., 2004). Second, applied studies by Lee, Ashton, and de Vries (2005) and Lee, Ashton, and Shin (2005) showed that the additional factor of H–H was at least as strongly related as any of the FFM dimensions to counterproductive behavior across five samples from four different countries.

TABLE 1

Study Descriptives and Intercorrelations by Country

	German sample			Canadian sample													
Variable	*M*	*SD*	α	*M*	*SD*	α	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. CWB	1.39	.51	.74	1.70	.77	.81		.54**	−.46**	−.23**	.07	−.03	−.31**	−.02	−.34**	−.47**	−.51**
2. CAB	2.96	.84	.88	2.75	.84	.89	.32**		−.45**	−.29**	.07	−.01	−.48**	−.05	−.50**	−.55**	−.67**
3. Honesty−Humility	3.50	.60	.83	3.23	.65	.86	−.38**	−.29**		.14	−.10	.31**	.24**	.15	.37**	.63**	.49**
4. Emotionality	3.30	.53	.83	3.59	.66	.86	−.13**	−.02	.06		.08	−.18*	.37**	−.17*	.16*	.35**	.25**
5. Extraversion	3.26	.45	.77	3.27	.68	.87	.06	.13*	−.20**	.11*		.04	−.01	.17*	−.21**	.08	−.20**
6. Agreeableness	3.02	.46	.78	2.88	.55	.80	−.10*	.02	.31**	−.15**	−.08		−.13	.02	.30**	.22**	.22**
7. Conscientiousness	3.47	.47	.82	3.74	.58	.85	−.35**	−.49**	.17**	.01	.06	.07		.04	.55**	.34**	.56**
8. Openness to Experience	3.39	.54	.78	3.39	.65	.82	.01	−.06	.09	.09	.24**	.04	.04		−.11	.06	.00
9. German personality-based integrity	3.34	.32	.85	3.27	.32	.80	−.41**	−.47**	.48**	−.03	−.22**	.48**	.54**	−.04		.55**	.74**
10. German overt integrity	3.44	.43	.92	3.56	.45	.91	−.54**	−.46**	.61**	.12**	.05	.30**	.37**	.14**	.59**		.57**
11. US personality-based integrity (CPI-Cp)	1.63	.14	.89	1.60	.14	.88	−.44**	−.67**	.49**	.01	−.27**	.24**	.43**	−.04	.73**	.65**	
12. U.S. overt integrity (EII)	3.36	.40	.91				−.62**	−.40**	.63**	.11	.00	.25**	.43**	−.01	.51**	.82**	–

Note. CWB = Counterproductive work behavior; CAB = Counterproductive academic behavior. Correlations for the German sample are given below the diagonal, for the Canadian sample above the diagonal. EII was not administered in the Canadian sample, and also not in the same German subsample who took the CPI-Cp. *N* between 169 and 171 in the Canadian sample. In the German sample, mean, *SD*, and correlations are based on the *N* corresponding to regression analyses reported in Table 2 (i.e., *N* for variables 1. and 3.–10. is 496–499, for variable 2. 259, for variable 11. 294, and for variable 12. 204.

*$p < .05$; **$p < .01$.

Moreover, H–H was more substantially related to an integrity test in two Canadian samples (Ashton, Lee, Morrison, Cordery, & Dunlop, 2006; Lee, Ashton, & de Vries, 2005). Thus, there is evidence that H–H bears more than just a semantic relationship to integrity tests. Consequently, Lee, Ashton, and de Vries and Lee, Ashton, and Shin suggested that H–H may be an important addition for understanding integrity test scores.

It is noteworthy, however, that Lee, Ashton, and de Vries (2005) limited this suggestion to the overt type of integrity tests, although there is currently no research on relations between the HEXACO model and personality-based integrity tests. The distinction between "overt" and "personality-based" integrity tests (Sackett, Burris, & Callahan, 1989) is now a commonplace framework for research on this topic. Roughly speaking, overt tests contain relatively transparent items often directly related to counterproductive 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.").

Because of these differences, personality-based integrity tests tend to be more broadly related to the domains covered in traditional personality inventories (see Marcus, 2000, for an overview), in which the FFM structure can often be recovered (e.g., Digman, 1990; but note that recovery is imperfect for the CPI, the most important predecessor of personality-based integrity tests, as found by McCrae, Costa, & Piedmont, 1993). Overt integrity tests, by contrast, have no immediate roots in the personality assessment tradition but were associated by leading test authors with theories of attitude–behavior relationships (Brooks & Arnold, 1989; Jones, 1991). Consistent with these different traditions, integrity test score–FFM relations were consistently, though not substantially, stronger for personality-based than for overt integrity tests (Marcus et al., 1997; Ones, 1993). Moreover, integrity test–FFM relations tend to rise if personality is measured with inventories originally designed to measure FFM dimensions (Marcus et al., 1997). In line with these meta-analytic findings, a head-to-head comparison of one overt and one personality-based integrity test as related to the NEO-PI-R (Costa & McCrae, 1992) version of the FFM found that a substantially larger proportion of personality-based than that of overt integrity test score variance was accounted for by a linear composite of FFM facets (adjusted $R^2 = .78$ vs. .49; Marcus et al., 2006).

None of the studies cited in the previous paragraph has included a measure of the H–H dimension, however. Recall that the only two studies to date that investigated relationships between integrity tests and the HEXACO model (Ashton et al., 2006; Lee, Ashton, & de Vries, 2005) found the most substantial correlations between integrity test scores and H–H but

looked at *overt* integrity tests only. Given these findings, combined with the evidence of comparatively weak relationships between traditional FFM dimensions and overt integrity tests, and the moral-evaluative nature of the H–H construct, which corresponds to the content of overt integrity tests tapping into evaluations of morally questionable behaviors, it seems plausible to assume that H–H will be particularly relevant for this type of integrity tests. Despite the different research traditions from which H–H and overt integrity tests emerged, the strong emphasis on morality (adjectives like honest or sincere define H–H, attitudes toward dishonest behaviors define large parts of the content of overt integrity tests) appears to provide for a stronger conceptual link between H–H with overt integrity tests than for any FFM dimension. In the absence of evidence on personality-based tests as a standard of comparison, this must remain a speculative statement, however.

To summarize, previous research has shown that (a) some FFM dimensions are moderately related to counterproductive behaviors, (b) there is a similar pattern of modest relationships between FFM dimensions and integrity test scores, (c) these latter relationships tend to be stronger for personality-based than for overt integrity tests, and (d) the relationship between H–H and overt integrity test scores tends to be stronger than for the Big Five. Taken together, these findings seem to indicate that, at the level of broad personality dimensions, personality-based integrity tests may derive their validity from personality factors other than H–H, whereas H–H is of additional relevance especially for overt integrity tests. However, this assumption is based on independent lines of research, which were never tested comprehensively in one data set. The major, but not the only, purpose of the present research is to fill this gap in the literature and tie the loose ends together. Our specific hypotheses and research strategy are outlined in the following section.

Present Study

Our first major objective in the present research was to address the issue of the sources of the relationships between integrity tests and counterproductive behavior within the personality sphere as described by the FFM and HEXACO models of personality. Whereas current knowledge on this issue is mostly based on indirect and partial evidence—linking either integrity tests to personality or to counterproductive behavior, or personality to the criterion—the present study included measures of counterproductive behavior, personality, and both types of integrity tests simultaneously, and thus allows for more direct tests of the links between these domains.

As our second major objective, we echoed the call of various authors (e.g., Marcus, Schuler, Quell, & Hümpfner, 2002; Ones & Viswesvaran, 2001) for more international research on integrity testing. There were few

studies outside the English speaking world, and almost all of these were conducted in just one country (Gough, 1971; Marcus, 2000, 2006a; Marcus et al., 2002, 2006; Salgado, Moscoso, & Lado, 2001) and thus provide only indirect evidence of convergence with North American results. To the best of our knowledge, the only cross-cultural study to date was that conducted by Fortmann, Leslie, and Cunningham (2002), who administered just one overt integrity test in English- and Spanish-speaking countries. In this research, we applied two measures of both overt and personality-based simultaneously to samples in Germany and Anglophone Canada, which permits a more comprehensive examination of cross-cultural equivalence than that previously available.

Having stated the various ways in which the present study goes beyond previous research, it may also be worth noting what this paper has *not* set out to accomplish. We were primarily interested in advancing the understanding of integrity test scores and do not explicitly address the incremental validity of H–H beyond FFM dimensions in predicting counterproductive behavior, although interested readers may interpret some of the evidence presented herein in that light. The incremental validity of H–H was the core issue of previous research using different samples (Lee, Ashton, & de Vries, 2005; Lee, Ashton, & Shin, 2005) and shall not be further discussed here. We also did not address the potential role of variables at higher or lower levels of the personality hierarchy, nor did we include conceptually distinct constructs like attitudes in our study. As mentioned at various places in the introduction, we do acknowledge the potential contribution of these additional variables but wanted to stay with a set of explanatory variables at one coherent conceptual level—dimensions of personality of the breadth defined by both the FFM and the HEXACO model. Instead of introducing further conceptual complexity, we preferred to rule out potential biases by employing multiple measures, settings, and samples for our study.

All in all, five different subsamples contributed to the database (cf. Methods section). In addition to being divided into Canadian and German participants, samples included employees as well as students, most but not all of whom were work experienced. As a consequence, we administered criterion measures of both counterproductive work behaviors (CWB), and counterproductive academic behaviors (CAB). This allowed us to examine the extent to which findings are affected by setting. Although cheating, plagiarism, or other types of CAB may not be directly relevant for industrial–organizational applications (but for another applied setting, of course), student self-reports of CAB (e.g., Hakstian et al., 2002; Woolley & Hakstian, 1993) and objective measures of student misbehavior (e.g., Cunningham, Wong, & Barbee, 1994; Mikulay & Goffin, 1998) have been repeatedly used to draw conclusions on the validity of integrity tests.

Furthermore, we administered two different overt and two different personality-based measures of integrity to the entire sample or at least to subgroups. Within each category, one integrity test was originally developed in North America and in the English language, and one was originally constructed in Germany and in the German language. This strategy not only allowed for a separation of possible cultural influences due to sample and test construction, respectively, but also permitted the first direct assessment of the convergence of integrity tests developed in different countries and languages.

With regard to our core research question of the sources of integrity tests' criterion-related validity in the sphere of personality, our research strategy was based on the following considerations. As outlined earlier, we expected personality sources of validity to differ between the two types of integrity tests. For personality-based integrity tests, the major source of validity can be assumed to lie within the five-factorial space described by the FFM. Although there is evidence that C, A, and ES stand out among the five factors in that respect, some findings (e.g., Wanek et al., 2003) suggest that a full account of the FFM may be most appropriate to test this supposition. Including all five factors also allowed for a more conservative test of our hypothesis on overt integrity tests. For overt integrity measures, we expected that the additional factor of H–H, as specified in the HEXACO model, accounts for a nontrivial proportion of integrity–criterion covariance beyond the Big Five.

To test these expectations, we applied a hierarchical multiple regressions approach. If our hypotheses hold, one could expect the following pattern of findings: If counterproductive behavior is regressed on integrity tests in the first step, we expected *relatively* larger incremental validity from adding the Big Five beyond overt than beyond personality-based integrity tests in the next step, as the criterion-relevant variance of FFM dimensions is better represented in personality-based than in overt tests. Adding H–H should yield the opposite results because H–H overlaps mostly with overt integrity tests. The pattern of results we expected is stated in the following hypotheses:

Hypothesis 1a: FFM factors add more to the prediction of counterproductive behavior beyond overt integrity tests than they add beyond personality-based integrity tests.

Hypothesis 1b: H–H adds more to the prediction of counterproductive behavior beyond personality-based integrity tests than it adds beyond overt integrity tests.

Largely equivalent hypotheses could be formulated using the opposite sequence of steps in hierarchical regressions (i.e., that overt tests add more beyond the FFM than beyond H–H, and vice versa for

personality-based tests). We entered integrity tests first, however, because this allows for a more direct comparison of incremental validities of FFM and H–H between the two types of integrity tests. We did not derive any specific expectations on possible effects of integrity tests' original source or language (German vs. English), country where data are collected (Germany vs. Canada), population (employees vs. students), or criterion (CWB vs. CAB). However, we address these issues in an exploratory fashion by reporting results separately for each of these variables.

Method

Samples and Procedures

Data were collected in two different places in Germany and in Canada. Canadian participants were $N = 171$ first year undergraduate students who received course credit in their introductory psychology course for participating in this study. Mean age in this subsample was 21.5 years ($SD = 4.7$); 26% were men. All but one Canadian participant reported to have current or prior work experience. Thus, we report results on both CWB and CAB for the (nearly) full Canadian sample.

Data collection in Germany took place by means of a snowball procedure. Undergraduate students received study participation credit for their own participation and additional credits for recruiting currently employed persons in addition, who completed the same set of measures (except for the CAB measure). Students in Germany have to obtain their required study participation credits during their first and second year of studies, but courses in I–O psychology and assessment for selection are not offered before the third year of studies. Hence, it can be assumed that participants were not trained in methods and topics relevant for the present research.

In order to avoid bias due to cheating (e.g., students filling in the employee questionnaires by themselves), a number of measures were taken. A coding system was developed that allowed to match student questionnaires with employee questionnaires submitted by the same student. Because some questions (e.g., age, number of hours worked) had an open response format, we were able to check the handwriting for (dis)similarity between the questionnaires submitted by each student. To ensure anonymity, each questionnaire was accompanied by a separate envelope, and employee participants were instructed to hand in the completed questionnaire to the student after sealing the envelope. Further, course credit for the employee survey was cut to the half amount to make cheating less attractive, and the submission of employee surveys was limited to two questionnaires per student. In addition, all student participants were explicitly instructed

to recruit permanently employed participants only, and employees were asked a number of questions about their current job. All questionnaires were checked by the first author upon submission, and there was no incidence of violations of these rules.

This procedure was employed consecutively in two waves, one of which took place in the former western and eastern parts of Germany, respectively. The East German subsample consisted of $N = 209$ employees and 114 students; 64 of the students reported having prior work experience. Mean age in the East German group was 31.8 years ($SD = 11.6$), and the proportion of men was 35%. Of the West German participants, $N = 215$ were employees and 144 were undergraduate students, and 81 persons within the latter subgroup had prior work experience and were administered both measures of CWB and CAB. Due to an error in editing the questionnaire, data on age were inadvertently not collected in West Germany. Thirty-seven percent of West German participants were men. Across the two German subgroups, there were 424 employees, 145 students with work experience, and 113 students without work experience. Forty-six percent of the employees and 19% of the students were men. As far as data on age were available, means were 36.8 years in employees ($SD = 11.2$), and 22.8 years in students ($SD = 5$).

Overall, there were $N = 853$ participants in this study, divided roughly equally between current employees and students. For the entire sample, mean age was 28.3 years ($SD = 10.9$), and the proportion of men was 34%.

Instruments

Personality. All participants took, at a minimum, the shortened, 96-item form of the *HEXACO-Personality Inventory* (HEXACO-PI, Lee & Ashton, 2004, 2006), the standard measure of personality as defined by the HEXACO model. This half-length version of the full HEXACO-PI does allow the assessment of 24 narrow facet scales making up the six higher-order factors, but its use is recommended primarily for the measurement of the six higher-order factors, which were at the core of the present investigation. Items are scored on a five-point response scale (1 = *strongly disagree* to 5 = *strongly agree*). The inventory was translated into German by fluently bilingual persons and subsequently reviewed by two of the authors to ensure accuracy of content. Previous studies with English and other language versions of the HEXACO-PI have generally found satisfactory psychometric properties, including reasonably high internal-consistency reliabilities, modest correlations between domain-level scales, and a well-recovered facet-level factor structure.

It is important to note that the last five dimensions of the HEXACO-PI form a space similar to that of the Big Five/FFM axes (see Lee & Ashton, 2004, 2006, for details). Specifically, the Extraversion, Conscientiousness, and Openness to Experience factors are nearly identical across the two models, whereas HEXACO-PI Agreeableness and Emotionality correspond roughly to rotated variants of Big Five/FFM Agreeableness and Emotional Stability (see Lee & Ashton, 2004, Table 6; Lee, Ogunfowora, & Ashton, 2005; Table 2). Therefore, the ability of these five HEXACO factors (i.e., other than Honesty–Humility) to predict a given variable is likely to be very similar to that of the Big Five/FFM (for previous, cross-culturally replicated evidence of this in the context of CWB and integrity, see Lee, Ashton, & de Vries, 2005). In the present research, therefore, we approximate the Big Five/FFM using the HEXACO variants of those factors.

Integrity. All participants completed the *Inventar Berufsbezogener Einstellungen und Selbsteinschätzungen* (IBES, Marcus, 2006a; parts of the empirical work on the IBES are published in the English language, see Marcus, 2003, 2006b; Marcus & Schuler, 2004; Marcus et al., 2002, 2006), an integrity inventory originally developed in German, which contains a 60-item overt and a 55-item personality-based part. The major objective in constructing the IBES was to design an inventory, which resembles the prototypical content themes repeatedly discovered in leading U.S. integrity tests of both types (e.g., Wanek, 1995). Except for an initial punitiveness subscale in the overt part, most of these themes could be clearly identified in the German data (e.g., rationalizations of deviant behavior, perceived CWB norms, behavioral intentions in the overt part; stimulus seeking, trouble avoidance, self-esteem in the personality-based part; cf. Marcus, 2006a; Marcus et al., 2006). One theme intentionally excluded from the IBES, however, is that of admission scales typically found in—especially overt—integrity tests, in order to avoid content overlap with self-reports of counterproductive behavior. Aside from these exceptions, the length, content, and breadth of measurement of this test correspond closely to those of many commercial integrity tests available in North America.

Items on the IBES are scored on a five-point Likert-type scale of endorsement. Previous studies using the IBES showed evidence of construct and criterion-related validity closely resembling meta-analytic findings on integrity tests in North America. Among other things, the two parts of the IBES were found to correlate with CWB, job performance ratings, job satisfaction, broad and narrow personality variables, cognitive ability, and demographic variables in similar ways as long established for American integrity tests (Marcus, 2006a; see also the English publications cited above). However, the present study is the first to test the IBES directly

TABLE 2

Hierarchical Regressions of CWB and CAB on Controls, Integrity Tests, and Personality

Criterion	CWB				CAB			
Integrity test	German		American		German		American	
	Overt	Pers-b	Overt	Pers-b	Overt	Pers-b	Overt	Pers-b
Step 1: Controls[1]								
Gender	.244***	.244***	.325***	.165***	.129**	.132**	−.040	.170**
Country	.209***	.209***	na	na	−.137**	−.134**	na	na
Population	−.067	−.067	na	na	na	na	na	na
R	.327***	.327***	.325***	.165***	.181**	.181**	.040	.170**
R^2	.107	.107	.106	.027	.033	.033	.002	.029
Step 2: Integrity								
Gender	.155***	.213***	.170**	.139**	.074	.106*	−.016	.038
Country	.302***	.247***	na	na	−.033	−.091*	na	na
Population	−.002	.031	na	na	na	na	na	na
Integrity	−.473***	−.361***	−.575***	−.460***	−.489***	−.470***	−.392***	−.665***
R	.565***	.481***	.642***	.488***	.509***	.501***	.393***	.673***
R^2	.319	.231	.412	.238	.259	.251	.155	.453
Adj. ΔR^2 (as compared to Step 1)	.212***	.125***	.305***	.210***	.226***	.218***	.147***	.423***
Effect size (f^2)	.313	.163	.522	.277	.306	.292	.183	.775
Step 3a: HEXACO FFM								
Gender	.163***	.190***	.185**	.126**	.092*	.084	.029	.051
Country	.370***	.301***	na	na	−.054	−.028	na	na
Population	.036	.029	na	na	na	na	na	na
Integrity	−.439***	−.283***	−.456***	−.476***	−.411***	−.392***	−.312**	−.569***
Emotionality	−.014	−.044	−.005	−.016	.074	−.002	.041	.039

(*continues*)

TABLE 2
(*continued*)

Criterion	CWB				CAB			
Integrity test	German		American		German		American	
	Overt	Pers-b	Overt	Pers-b	Overt	Pers-b	Overt	Pers-b
Extraversion	.102**	.039	.021	−.005	.121**	.023	.069	.005
Agreeableness	.048	.038	−.031	.038	.094*	.152**	−.047	.143**
Conscientiousness	−.172**	−.163***	−.248***	.013	−.351***	−.250***	−.216*	−.220***
Openness to Experience	.043	−.031	.034	−.041	−.024	−.081	−.094	.000
R	.600***	.506***	.683***	.491***	.624***	.580***	.464***	.718***
R^2	.360	.256	.466	.241	.389	.336	.215	.516
Adj.ΔR^2 (as compared to Step 2)	.036***	.019**	.041**	−.005	.124***	.078***	.023	.056***
Effect size (f^2)	.063	.034	.101	.004	.213	.128	.076	.130
Step 3b: HEXACO H−H								
Gender	.147***	.177***	.163**	.128**	.072	.095*	.029	.035
Country	.281***	.226***	na	na	−.066	−.124**	na	na
Population	.021	.067	na	na	na	na	na	na
Integrity	−.392***	−.249***	−.498***	−.353***	−.416***	−.396***	−.301**	−.649***
Honesty–Humility	−.139***	−.267***	−.124	−.218***	−.136**	−.212***	−.277**	−.038
R	.575***	.534***	.649***	.523***	.522***	.538***	.470***	.674***
R^2	.331	.285	.421	.274	.272	.290	.221	.454
Adj. ΔR^2 (as compared to Step 1)	.011**	.052***	.007	.034***	.011**	.037***	.059**	.000
Effect size (f^2)	.016	.074	.016	.049	.018	.054	.084	.002
Relative increment of H–H^2	.058 (.087)		.073 (.130)		.072 (.121)		−.092 (−.136)	

(*continues*)

TABLE 2
(*continued*)

Criterion	CWB				CAB			
Integrity test	German		American		German		American	
	Overt	Pers-b	Overt	Pers-b	Overt	Pers-b	Overt	Pers-b
Step 3c: HEXACO all								
Gender	.150***	.154***	.172**	.119*	.092*	.084	.069	.048
Country	.350***	.286***	na	na	.022	−.064	na	na
Population	.059	.079	na	na	na	na	na	na
Integrity	−.359***	−.182***	−.386***	−.360***	−.345***	−.332***	−.259**	−.518***
Emotionality	.013	−.024	−.006	.003	.082	.035	.077	.053
Extraversion	.073*	.003	−.004	−.025	.099*	.002	.040	−.002
Agreeableness	.069*	.087*	−.004	.083	.117**	.197***	−.015	.169***
Conscientiousness	−.174***	−.168***	−.246***	−.016	−.348***	−.244***	−.172	−.237***
Openness to Experience	.056	.035	.045	.005	−.014	−.035	−.061	.000
Honesty–Humility	−.145**	−.298***	−.129	−.243***	−.137**	−.240***	−.233*	−.121*
R	.609***	.564***	.689***	.530***	.633***	.618***	.506***	.725***
R^2	.371	.318	.474	.281	.401	.382	.256	.526
N	663	663	204	460	425	423	114	310

[1]Controls are dummy-coded as follows: Gender: 0 = *female*, 1 = *male*; country: 0 = *Germany*, 1 = *Canada*; population: 0 = *student*, 1 = *employee*.
[2]Relative increment of H–H is computed by subtracting, in terms of effect size-adjusted ΔR^2 (f^2 in parentheses), the difference of H–H (2b) vs. FFM (2a) increments over personality-based integrity from adjacent column H–H vs. FFM increments over overt integrity. Entries corresponding to single variables are standardized regression coefficients (βs). Effect sizes used for hypotheses testing are given boldface. Significance levels for incremental validity are based on ΔF. Pers-b: personality-based; adj. ΔR^2: adjusted ΔR^2.
$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

vis-à-vis American integrity tests. The IBES was translated into English with a similar procedure as the HEXACO-PI. A fluently bilingual English native speaker prepared the original translation, which was then carefully discussed between two of the present authors, one of whom is a bilingual native German speaker and one is a native English speaker, until consensus was reached.

In addition to the IBES, participants in the Canadian and the West German subsamples completed a publicly available personality-based integrity test, which is based on the content of the *California Psychological Inventory* (CPI-Cp, Hakstian et al., 2002). The CPI, in particular its socialization scale, is the progenitor of the entire group of personality-based integrity tests (Gough, 1971). The CPI-Cp scale is a collection of 80 CPI items from various subscales specifically designed to predict counterproductive behavior. It was found to show expected relationships with personality, CWB, and CAB in a series of validation studies (Hakstian et al., 2002). Items have a dichotomous true–false response format. After permission of the test publisher was received, English items were translated by two bilingual native German speakers working independently and subsequently reviewed by the first author. If necessary, disagreements were discussed among translators and the first author until they could be resolved. This procedure was used with all remaining English-to-German translations and shall not be reiterated below.

The East German participants did not complete the CPI-Cp but instead took a public-domain overt integrity test, the *Employee Integrity Inventory* (EII, Ryan & Sackett, 1987). The content of the EII was modeled after leading commercial overt integrity tests; it is widely used in noncommercial integrity test research (a SSCI search retrieved 43 citations). The inventory is composed of two parts, a 52-item theft attitudes scale with a Likert-type response format and an 11-item admission scale whose items are scored on frequency scales. For the present research, we applied only the EII attitudes scale, in order to avoid content overlap with measures of counterproductive behavior.

Counterproductive behavior. We measured CWB in all employees and those Canadian and West-German students who reported work experience of at least 1 month within the previous 3 years (in the small subsample of work-experienced East-German students, we dropped the CWB measure, to avoid an overly long administration time due to some additional measures applied there). For that purpose, we applied the *Workplace Behavior Questionnaire* (WBQ, Ashton, 1998), which contains eight scored items tapping into various kinds of CWB, including theft, absenteeism, vandalism, and alcohol use. Items are scored on varying scales, the format of which depends on the type of CWB (e.g., percentage of work shifts missed, value in dollar—or euro—of stolen items). Evidence of

convergent validity of the WBQ with other measures of CWB is reviewed by Lee, Ashton, and de Vries (2005), who report a disattenuated correlation of .86 of the WBQ with another self-report measure of CWB, which itself was substantially correlated with observed theft in a laboratory setting.

CAB was measured in students only, using the *Inventory of Counterproductive Behavior* (ICB, Hakstian et al., 2002, who report a series of studies establishing the measure's factorial validity as well as cross-validated correlations with integrity and personality tests closely resembling findings for other self-report measures of counterproductive behaviors). The ICB was developed for use with university students but, in its full 40-item version, contains a number of items related to work or not unequivocally tapping into academic misconduct. As we wanted to separate CWB clearly from CAB, we dropped all nonacademic items as identified in a procedure involving ratings and discussions. Finally, 26 items with unequivocally academic content were retained for the measure. Of the nine ICB factor-analytically derived subscales reported in Hakstian et al. (2002), the components of cheating (e.g., looking at a classmate's exam), low personal standards (e.g. turning in work of unnecessarily poor quality), duplicity (e.g., false claims of medical illness), and misrepresentation (e.g., plagiarism on assignments) were completely retained. In the remaining scales, one or more items were dropped, and the short nonacademic scales of property theft and work avoidance were eliminated entirely. Items in the ICB are scored on 6-point frequency scales (*not even considered it* to *did it three or more times*).

Results

Bivariate Relationships and Cross-Cultural Equivalence

Table 1 shows means, standard deviations, internal consistency reliabilities, and intercorrelations of all study variables for the Canadian and German samples separately. It is evident that all integrity tests showed consistently substantial relationships with both criteria. Of the personality dimensions, H–H and C were almost as highly related to criteria as were integrity tests, whereas bivariate criterion-related validities of the remaining factors ranged from negligibly small to moderate (in case of Emotionality in the Canadian sample) at best. However, H–H and overt integrity tests tended to be slightly more strongly correlated with work-related measures of counterproductive behavior rather than academic-based measures, whereas the opposite pattern emerged for C and personality-based integrity tests. Also noteworthy are correlations among integrity tests. The German and American personality-based test correlated uniformly highly in German ($r = .73$) and Canadian (.74) samples. Relationships with the

German overt test were somewhat lower but still substantial (*r*s range from .55 to .65). With the American overt test (only measured in the East German sample), the gap between convergent and divergent correlations with other integrity measures was larger. The two overt integrity measures correlated at a very high *r* of .82, whereas the American overt and German personality-based integrity tests correlated at $r = .51$. Overall, these findings show a high degree of convergence between German and American integrity tests across countries, and the pattern of convergent and divergent relations supports the validity of the distinction between types of integrity tests.

In addition to bivariate analyses, we performed a simple omnibus test of structural equivalence between the two countries using multisample structural equations modeling. Imposing equality constraints on correlations between those variables that were uniformly measured in the Canadian and German samples (the six personality dimensions and German integrity tests) supported the equivalence of the two matrices. Although model χ^2 (61.3, $df = 36$, $p = .005$) was statistically significant with this relatively large sample, the χ^2/df ratio (1.7) as well as other applicable fit indices (RMSEA = .037, NFI = .96, NNFI = .98, CFI = .99, IFI = .99) indicated good to excellent fit (cf. Hu & Bentler, 1999). Separate tests of structural equivalence for the two former parts of Germany indicated a closer fit of correlation matrices between Canadian and West German samples ($\chi^2_{(36)} = 56.5$; RMSEA = .042, remaining indices between .95 and .98) than between East and West German samples ($\chi^2_{(36)} = 99.3$; RMSEA = .067, other indices between .93 and .95). Because multisample comparisons indicated acceptable to good fit for the assumption of equivalence, on one hand, but some variation in two-culture comparisons, on the other, we report tests of our hypotheses for both data collapsed across countries (Table 2) and for Canadian, West German, and East German participants separately (Table 3).

Criterion-Related Validity Comparisons across Integrity Tests

As outlined previously (cf. "Present Study"), we employed a hierarchical multiple regressions approach to test our hypotheses on the different sources of integrity—counterproductivity relationships within the personality sphere across types of integrity tests. We expected that the Big Five factors would add more to the prediction of criteria beyond the overt than beyond the personality-based test, and we expected to find the opposite pattern for H–H. In each analysis, demographic controls (as far as available) were entered first, followed by integrity at Step 2, followed by the Big Five at Step 3a, or, alternatively, H–H at Step 3b. We also report data on analyses with all personality dimensions entered in the final step (3c).

TABLE 3

Hierarchical Regressions of Criteria on Gender, Integrity, and Personality by Region and Population

Sample	East German Empl.		West German Empl.		East German Stud.		West German. Stud.		Canadian Students			
Criterion	CWB		CWB		CAB		CAB		CWB		CAB	
Integrity test	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb
Step 1: Gender[1]												
Gender (β equals R)	.325***	.325***	.149*	.145*	−.040	−.040	.088	.095	.280**	.280**	.270***	.270***
R^2	.106	.106	.022	.021	.002	.002	.008	.009	.078	.078	.073	.073
Step 2: Integrity												
Gender	.191**	.283***	.085	.131*	.020	−.008	.005	.088	.188**	.232**	.158*	.196**
Integrity	−.536***	−.418***	−.510***	−.410***	−.421***	−.488***	−.512***	−.438***	−.430***	−.301***	−.514***	−.463***
R	.613***	.528***	.527***	.435***	.419***	.488***	.513***	.448***	.505***	.408***	.569***	.531***
R^2	.375	.279	.278	.189	.175	.238	.263	.201	.255	.167	.324	.282
Adj. ΔR^2 (as compared to Step 1)	.268***	.171***	.254***	.165***	.167***	.232***	.251***	.187***	.173***	.084***	.249***	.207***
Effect size (f^2)	.432	.240	.355	.207	.211	.311	.346	.240	.238	.106	.371	.291
Step 3a: HEXACO FFM												
Gender	.195**	.277***	.102	.107	.059	−.004	.160*	.204**	.182*	.174*	.095	.088
Integrity	−.464***	−.232*	−.493***	−.385***	−.368***	−.593***	−.351***	−.298**	−.427***	−.253*	−.453**	−.386***
Emotionality	−.010	−.012	−.002	−.061	.060	.043	.162*	.134	.057	−.080	.024	−.120
Extraversion	.061	.005	.176**	.100	.094	−.017	.142*	.038	.093	.037	.105	.017
Agreeableness	.043	−.015	.092	.106	.041	.160	.186*	.279**	.040	.011	.050	.062
Conscientiousness	−.266***	−.301***	−.132*	−.093	−.231*	.023	−.456***	−.448***	−.120	−.083	−.298***	−.187**

(*continues*)

TABLE 3
(*continued*)

Sample	East German Empl.		West German Empl.		East German Stud.		West German. Stud.		Canadian Students			
Criterion	CWB		CWB		CAB		CAB		CWB		CAB	
Integrity test	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb
Openness to Experience	.078	.034	.045	−.032	−.082	−.137	−.036	−.066	−.019	−.056	−.020	−.105
R	.675***	.582***	.574***	.467***	.489***	.521***	.706***	.682***	.529***	.426***	.646***	.588***
R^2	.455	.339	.329	.218	.240	.272	.498	.460	.280	.181	.418	.345
Adj. ΔR^2 (as compared to Step 2)	.067***	.038**	.035*	.010	.029	−.001	.220***	.244***	.005	−.011	.077***	.043*
Effect size (f^2)	.147	.091	.076	.037	.084	.045	.468	.485	.047	.018	.162	.096
Step 3b: HEXACO H–H												
Gender	.183**	.236***	.085	.116	.051	.024	.010	.072	.199**	.218**	.166*	.183**
Integrity	−.445***	−.275***	−.509***	−.344***	−.321**	−.402***	−.524***	−.418***	−.253**	−.165*	−.393***	−.353***
Honesty–Humility	−.141	−.277***	−.001	−.140	−.248**	−.210*	.028	−.070	−.279**	−.375***	−.188*	−.304***
R	.622***	.576***	.527***	.452***	.474***	.523***	.513***	.453***	.550***	.537***	.588***	.602***
R^2	.387	.331	.278	.204	.225	.274	.263	.187	.302	.288	.345	.362
Adj. ΔR^2 (as compared to Step 2)	.008	.049***	−.004	.011	.044**	.027*	−.005	−.002	.044**	.112***	.018*	.076***
Effect size (f^2)	.018	.078	.000	.019	.065	.048	.001	.005	.067	.171	.032	.125
Relative increment of H–H^2	.070 (.116)		.040 (.058)		.013 (.022)		−.019 (−.013)		.084 (.121)		.092 (.159)	

(*continues*)

TABLE 3
continued

Sample	East German Empl.		West German Empl.		East German Stud.		West German. Stud.		Canadian Students			
Criterion	CWB		CWB		CAB		CAB		CWB		CAB	
Integrity test	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb	IBES-ov	IBES-pb
Step 3c: HEXACO all												
Gender	.182**	.215**	.105	.092	.088	.035	.154*	.178*	.194*	.194*	.101	.102
Integrity	−.384***	−.100	−.509***	−.322**	−.306**	−.521***	−.320***	−.260**	−.258**	−.171	−.351***	−.320***
Emotionality	−.011	−.014	−.002	−.052	.087	.075	.163*	.146	.059	−.001	.025	−.057
Extraversion	.028	−.046	.182**	.081	.065	−.032	.133*	.031	.040	−.017	.075	−.020
Agreeableness	.057	.026	.089	.124	.053	.163	.198**	.300***	.098	.129	.085	.157*
Conscientiousness	−.263***	−.299***	−.132*	−.101	−.192*	−.035	−.465***	−.460***	−.098	−.039	−.286***	−.156
Openness to Experience	.083	.062	.042	−.006	−.055	−.103	−.025	−.037	.062	.040	.004	−.036
Honesty–Humility	−.132	−.316***	.029	−.149*	−.204*	−.205*	−.064	−.152*	−.297**	−.413***	−.176*	−.324***
R	.681***	.635***	.574***	.484***	.519***	.551***	.708***	.693***	.571***	.555***	.659***	.651***
R^2	.464	.403	.329	.234	.270	.303	.501	.481	.327	.308	.434	.424
N	204	204	209	210	114	114	141	139	169	169	170	170

[1]Gender is dummy-coded as 0 = *female*, 1 = *male*.

[2]See Table 2 for computation of relative increment of H–H. Entries corresponding to single variables are standardized regression coefficients (βs). Effect sizes used for hypotheses testing are given boldface. Significance levels for incremental validity are based ΔF. Empl. = employees; Stud. = students; IBES-ov: overt German integrity test; IBES-pb: personality-based German integrity test; adj. ΔR^2: adjusted ΔR^2.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 2 shows the results of these analyses for each integrity test and each criterion separately, but collapsed across samples. We report three different types of effect sizes: adjusted ΔR^2 and f^2, and an index labeled "relative increment of H–H" explained below. Adjusted ΔR^2 takes into account possible effects of differences in sample size and number of predictors across analyses, which is important because we partially compare the single construct of H–H with a combination of the Big Five. Cohen's (1988) effect size measure $f^2(\Delta R^2/(1 - R^2))$, on the other hand, controls for possible biases in R^2 attributable to the fact that incremental validities in hierarchical regression depend in part on how much criterion variance was already explained in previous steps, and thus on the order in which variables are entered.

Neither adjusted ΔR^2 nor f^2 account for possible biases due to general differences in incremental validities between FFM dimensions and H–H. Suppose that the FFM (H–H) is, on average, more valid beyond integrity tests than H–H (FFM). In this situation, one may find no support for Hypotheses 1b (1a) and, at the same time, overestimate support for Hypothesis 1a (1b), given the overall expected pattern of differences holds. To control for this possible effect, we computed a relative index based on sets of four effect sizes, which is computed in the following way: We took the difference between the incremental effect sizes of H–H and FFM dimensions beyond the personality-based integrity test and, from that result, we subtracted the corresponding difference with regard to the overt integrity test in the adjacent column to the left. For example, in the first comparison reported in Table 2, we computed the difference between adjusted ΔR^2 of Models 3b and 3a in the second data column (.052 − .019) and subtracted the corresponding difference in the first data column (.011 − .036), which gives an overall relative increment of .058 (i.e., .033 – (−.025)). If our hypotheses hold, the resulting overall increment should become positive. Effectively, this measure estimates the joint effects of the differences between types of integrity tests specified in our hypotheses, while holding average differences between FFM dimensions and H–H constant.

In the following, we base our tests Hypotheses 1a (FFM dimensions add more beyond overt than beyond personality-based integrity test scores) and 1b (H–H adds more beyond personality-based than beyond overt integrity test scores) largely on the most conservative estimates of adjusted ΔR^2. That is, we compare the adjusted ΔR^2 of FFM dimensions beyond overt and personality-based integrity tests in adjacent columns to test Hypothesis 1a and do the same for H–H to test Hypothesis 1b. The remaining effect sizes are reported cursorily in the tables because of their distinctive features. Notably, inspection of these additional effect sizes does not change any of our conclusions but consistently show stronger effects than

adjusted ΔR^2. We scaled all outcomes such that a positive sign indicates support for our hypotheses.

As results in Table 2 are in part based on different (but partially overlapping) samples and on integrity tests developed in different countries and languages, the comparability of findings varies within the table. Most directly comparable are results between the German overt and personality-based integrity tests because these are based on the same samples and the same origin of the integrity tests. The two American integrity tests share the same cultural origin, but the samples do not overlap. We focus on these two comparisons (German overt vs. personality-based, and American overt vs. personality-based) in order to keep the number of alternative explanations at a minimum.

After entering gender, and, as far as applicable, country and population (student vs. employee) as controls (age was measured only in a subsample and had trivial effects in this group; details are available from the first author), the expected pattern as specified in Hypotheses 1a and 1b emerged in all cases but one. Specifically, the Big Five added more to the prediction of both work and academic counterproductive behavior beyond the overt than beyond the personality-based German integrity test (adjusted $\Delta R^2 = .036$ vs. .019, and .124 vs. .078, respectively), whereas H–H had lower incremental validity above the overt than above the personality-based integrity measure for both work (.011 vs. .052) and academic counterproductive behaviors (.011 vs. .037). Similarly, the Big Five accounted for more variance in the work counterproductive behavior criterion beyond the overt American integrity test than beyond the personality-based measure (.041 vs. −.005), and this pattern was reversed when H–H was added instead of FFM dimensions (.007 vs. .034). In contrast, patterns contrary to our hypotheses emerged only with the American integrity tests predicting CAB. The adjusted ΔR^2 of the FFM dimensions was larger beyond personality-based (.056) than beyond overt integrity (.023), but the reverse pattern was observed for H–H (.000 vs. .059). Averaged across comparisons, however, both Hypothesis 1a and 1b were supported. The mean adjusted ΔR^2 of FFM dimensions was larger beyond overt than beyond personality-based integrity tests (unweighted absolute difference = .019, or 51%; weighted = .029, or 88%), and the opposite was found when H–H was added to integrity tests (unweighted = .009, or 41%; weighted = .022, or 157%), especially when sample size was taken into account.

Beyond these findings pertaining to the expected *pattern* of results, it needs to be mentioned that f^2 effect sizes for all predictors varied considerably across analyses. The f^2 for integrity tests ranges from .16 to .78, and respective increments of the personality dimensions beyond integrity tests cover a range from conventional levels (Cohen, 1988) of small ($f^2 = .02$) beyond that of medium-sized ($f^2 = .15$) effects. Further, it is evident

that the overlap between personality dimensions and integrity test scores, although it was sizeable (particularly between H–H and overt integrity), cannot fully account for the criterion-related validity of integrity tests. If all personality dimensions are added to the models, regression coefficients for integrity tests were substantially reduced with some consistency but remained sizeable and statistically significant in all eight models.

Comparisons Across Cultures and Populations

In addition to the multiple comparisons testing our hypotheses with the largest samples available in our data set, we were also interested in the extent to which our findings hold across the different samples from which data were collected. Although we controlled for cultural and sociodemographic variables statistically as far as possible in the preceding analysis, direct head-to-head comparisons with identical subsamples composed of individuals with similar background provide for more direct tests of possible moderator effects of these variables. Of course, this comes at the price of substantially reduced sizes of each subsample. For the following analyses, we report results separately for Canadian students, and for West and East German students and employees, respectively. We divided the German sample because of the earlier reported indication that differences between East and West Germans may be even more pronounced than between Canadians and West Germans. For German students, we refrained from reporting results on work counterproductive behavior for the small group who took this measure in addition to the academic counterproductive criterion. Gender was again controlled statistically. As the American overt and personality-based integrity tests were administered in nonoverlapping subsamples, we report only results on the German integrity tests in this subsection. Otherwise, we followed the same procedures as with the full samples.

The results of hierarchical regression analyses comparing criterion-related validities across subsamples and criteria are shown in Table 3. Similar to the more general analyses reported above, results confirmed our Hypotheses 1a, and 1b, respectively, in all but one of the six comparisons. Adjusted ΔR^2 of FFM dimensions beyond overt versus personality-based integrity tests ranged from $-.024$ to .034 (both values were found in the same groups as with the relative increments), with similar unweighted (.018 or 33%) and weighted (.019 or 43%) means. For H–H, adjusted ΔR^2 beyond personality-based versus overt integrity tests ranged from $-.017$ (East German students, academic counterproductive behavior as criterion) to .068 (Canadian students, work counterproductivity as criterion). The mean across comparisons was .028 regardless of weighting. Although the mean absolute difference was not affected by weighting (.028 in both

cases), the mean increment of H–H was 156% larger beyond personality-based than beyond overt integrity tests in the unweighted case and even 200% after taking sample size into account.

With the exception of the Canadian participants, who were administered both measures of work and academic counterproductive behavior and thus are listed twice in Table 3, all analyses reported in this table are based on independent samples. Hence, unlike in the general analyses reported above, we can examine possible systematic differences between groups and variables by computing mean values across subsamples that may be interpreted as "bare bones" meta-analytic estimates, although admittedly based on very small-scale meta-analyses. In predicting work-related counterproductivity, the meta-analytic means of adjusted ΔR^2, weighted by sample size ($K = 3, N = 582$), were .024 for testing Hypothesis 1a and .040 for the test of Hypothesis 1b. The corresponding values for predicting academic counterproductive behavior ($K = 3$, $N = 424$) were comparatively smaller: .011 (Hypothesis 1a) and .020 (Hypothesis 1b). Effects were strongest in Canadians (averaged across criteria: .025, .063; same order as above), followed by East Germans ($K = 2$, $N = 318$; mean effects: .026, .020), followed by West Germans ($K = 2$, $N = 349$; mean effects: .005, .010). In addition, somewhat stronger mean effects were observed for employees ($K = 2$, $N = 413$; mean effects: .027, .028) than for students ($K = 3$, $N = 423$; mean effects: .007, .022; averaged within the Canadian sample). Hence, values vary across subsamples but all means in all groups lie in the expected direction. Due to the very limited number of samples and the unusual nature of our effect sizes (relative rather than absolute, based on squared multiple correlations; cf. Hunter & Schmidt, 2004), we refrain from reporting estimates of variation.

Discussion

The present study addressed the question as to what extent broad dimensions of personality contribute—and possibly add—to the established validity of integrity tests for predicting counterproductive behaviors. This research was more comprehensive than previous studies on the same issue in various respects. First, we investigated the three sets of variables involved—integrity, personality dimensions, and counterproductive behaviors—simultaneously instead of trying to draw conclusions based on various combinations of only two sets of variables. Second, we included in our research both a variant of the Big Five dimensions of personality, which are traditionally at the core of personality research in I–O psychology, and a sixth dimension recently found to be an important addition to the FFM in research on counterproductive behavior and integrity testing (Lee, Ashton, & de Vries, 2005; Lee, Ashton, & Shin, 2005). Third, we

distinguished between the two categories of integrity tests conceptually instead of collapsing them in deriving our hypotheses. Fourth, in testing our hypotheses, we examined two exemplars of integrity tests per category (with those exemplars developed in different countries) and two criteria in different settings, thereby allowing us to test the stability of results. Finally, this was the first study in which multiple integrity tests of both types were administered simultaneously in two different countries and languages, which allowed us to address a number of unresolved cross-cultural issues in this literature. Below, we discuss the implications of this research, followed by an account of its limitations.

Implications

In general, findings supported our expectation that the two types of integrity tests derive their criterion-related validity from *partially* different sources within the personality domain. We emphasize the word "partially" because our findings speak to the coexistence of similarities and differences between overt and personality-based integrity tests. Bivariate correlations reported in Table 1 show that integrity tests of both types were mostly related to the same personality dimensions, which is in line with previous research on the FFM and integrity tests (Marcus et al., 1997; Ones, 1993) and extends earlier findings on the HEXACO model as related to overt integrity tests (Lee, Ashton, & de Vries, 2005; Lee, Ashton, & Shin, 2005). Although there is evidence of similarity between overt and personality-based integrity tests, we also found some consistent differences in bivariate correlations reported in Table 1. Relationships were consistently stronger between integrity and C with personality-based than with overt integrity tests, whereas the opposite pattern emerged with respect to H–H. In addition, personality-based, but not overt, integrity tests showed moderately negative correlations with Extraversion.

Findings from hierarchical regression analyses confirm that both the overall similarity and the specific differences between types of integrity tests translate into relationships with behavioral criteria, which was the core issue of the present study. Consistent with bivariate findings showing similarity, the validity of both types of integrity tests is reduced by adding the Big Five as well as H–H, indicating that integrity tests of all kinds in part owe their criterion-related validity to these sources. Consistent with bivariate evidence of dissimilarity, the relative importance of the Big Five and H–H dimensions differed between types of integrity tests. These differences followed a systematic pattern, which was predicted based on conceptual considerations and previous findings: The Big Five were more important for explaining the criterion-related validity of personality-based integrity tests, whereas H–H was of greater significance for overt integrity

tests. This difference was small in terms of absolute effect sizes but appears sizable on a relative scale that takes into account that incremental validities of our personality measures beyond integrity were generally not large.

This pattern of findings has noteworthy implications for the study of integrity tests as the prototypical exemplar of the type of empirically derived multidimensional predictors referred to as COPS, as well as for the application of personality theories to work settings in general. With respect to integrity testing, the distinction between overt and personality-based integrity tests may require some theoretical refinement. Introduced by Sackett et al. (1989) more than 15 years ago, this distinction is widely applied in integrity test research and there is evidence for its usefulness in demarcating the presence, absence, and degree of differences between integrity tests in a wide range of important topics (e.g., Alliger & Dwight, 2000; Ones et al., 1993; Wanek et al., 2003). Most commercial integrity tests can be unequivocally assigned to one of these two categories.

Despite its proven utility, however, the distinction appears to be remarkably deficient at a conceptual level. As discussed in some detail by Marcus and colleagues (Marcus et al., 1997, 2006), the categories adequately describe different (and independently developed) traditions of integrity test construction, but as a pair of terms, "overt" and "personality-based" are neither mutually exclusive categories nor do they refer to the same level of abstraction. "Overt" refers to one feature of item content, whereas "personality-based" refers to an entire subdiscipline of psychology. Although the latter term defines the theoretical domain in which the meaning of integrity test scores can be sought, "overt" does not provide researchers with equivalent guidance. The bases of overt integrity tests' content and sources of validity within and beyond the personality domain are neither conceptually as well defined nor empirically equally well accounted for as is the case for personality-based integrity tests.

The present research leads to the conclusion that H–H appears to be at least as important for a theoretical understanding of overt integrity tests as any one of the Big Five or perhaps even a combination of these traits. H–H alone was more highly related to overt integrity measured via IBES in Germany ($r = .54$) and Canada ($r = .64$), and to the other overt measure (EII) in Germany ($r = .62$), than an optimally weighted combination of the Big Five in any of these cases ($R = .49, .53, .45$, respectively). This can be taken as evidence that H–H adds significantly to our understanding of overt integrity tests in particular. Still, there is also evidence that additional explanations are needed to account for the meaning of overt integrity test scores. If we regressed integrity test scores on all personality dimensions, we found a higher proportion of variance accounted for in personality-based than in overt integrity with both the German (56.2% vs. 43.1%) and

the American (44.7% vs. 37.8%) integrity tests. The differences by test type were less pronounced than that reported earlier (78% vs. 49%) in a study using the NEO version of the FFM (i.e., including FFM facets; Marcus et al., 2006), but the overall percentages were lower in the present study.

This pattern of findings implies that the inclusion of personality facets below the FFM domain level can add further to the understanding of integrity tests, but more so for the personality-based type, whereas the inclusion of H–H at the same level as FFM dimensions adds more to the understanding of overt than of personality-based integrity tests. One promising direction for future research hence may be sought in the inclusion of the facet level below the *six*-factorial space as described in the HEXACO model of personality. For example, the fairness facet of the HEXACO Honesty–Humility domain assesses a reluctance to cheat or to exploit others and, thus, is likely to be strongly related to overt integrity tests and to the criteria predicted by those tests. Another promising avenue for future research on overt integrity tests may be extensions beyond the sphere traditionally covered in models of personality. Based on the fact that leading overt integrity tests emerged from the tradition of attitude research and on evidence that attitudinal constructs are not adequately captured by the trait dimensions described in the Big Five model of personality (Paunonen & Jackson, 2000; Saucier & Goldberg, 1998), Marcus and colleagues (1997, 2006) proposed to seek a fuller account of overt integrity tests by adding attitudinal constructs to potential explanations for those measures' validity and even to replace the term "overt" with "attitudinal" (which would then be contrasted to "trait-based"). These extensions, however, were beyond the scope of the present paper, where we primarily tried to combine the tradition of relating integrity test scores to FFM dimensions with the recently emerged extension of the HEXACO model.

Along those lines of reasoning, the present findings may also have implications for broader issues in applications of personality theories and assessment in work settings. In the past 15 years, research in this area in general, and major reviews of the field in particular, relied on the FFM to an extent that made even a second-order meta-analysis possible (Barrick, Mount, & Judge, 2001). No doubt, the organization of personality constructs in the FFM has significantly contributed to the advancement of applied personality research. However, there were always scholars who warned that something important will be missing if personality is reduced to just five broad dimensions (e.g., Hough, 1992; Paunonen & Ashton, 2001; Schneider et al., 1996). Although we mentioned some evidence in the introduction that a trait even broader than the Big Five may provide for an extremely parsimonious explanation of relations between personality and work behaviors, the present results rather point to the necessity of additions to the FFM.

Although the differences between types of integrity tests and personality dimensions revealed in the present research imply mainly theoretical refinements, the consistency of these findings across single integrity tests, countries, and criteria also have practical implications. In their review of COPS, Ones and Viswesvaran (2001) identified the question as to whether North American findings on integrity testing generalize to other cultures and countries as a major avenue for future research in our increasingly globalized economic world. The present study addressed this question by collecting data in North America and Germany on integrity tests originally developed in North America and Germany. Our results point to a positive answer to Ones and Viswesvaran's question. Not only was the overall correlational structure of our study variables equivalent across countries, we also replicated the rather specific pattern of predictor–criterion relationships we had expected across different countries (or parts of one country) where data were collected and across different origins and languages of integrity tests. This is an encouraging finding for practitioners who wish to apply a predictor of job-related behaviors that has been found—in North America—to be more incrementally valid beyond cognitive ability than any other selection procedure (Schmidt & Hunter, 1998) in cultures where there was previously very little evidence of validity. It means that it may not be necessary to reinvent the wheel to arrive at similarly useful predictions.

Largely, our findings were also generalizable across work and academic settings, and whether we asked students or employees. In line with the few previous studies comparing work counterproductivity with academic counterproductive behavior (Hakstian et al. 2002; Woolley & Hakstian, 1993), personality correlates of academic counterproductivity did not differ fundamentally from those of the work-related criterion, and we also found no evidence that counterproductive behaviors have different causes in students compared to employees. Although these findings generally point to a high degree of parsimony in theoretical explanations across settings and populations, the slight differences between types of integrity tests and personality dimensions as related to work and academic counterproductive behavior are also notable. Unlike the work-related measure, our academic measure included content related to slow or sloppy work, which may be more closely related than unequivocally deviant behaviors to C and the C-driven part of personality-based integrity tests. As some measures of work-related counterproductive behavior include similar content while others do not, the observed difference may be more a function of varying definitions of the domain of counterproductive behaviors than of actual setting.

Study Limitations

Although the present study goes beyond previous research in several ways, it may nevertheless be criticized on various grounds. Perhaps the

most obvious cause for concern is the fact that all variables were measured with self-reports. Although it would have been desirable to validate our findings against nonself-report criteria, this was impractical in the present research. Our emphasis was on collecting data in a broad variety of settings to rule out the possibility that our findings were affected by variables such as culture, language, type of population, or type of counterproductive behavior. To be able to rule out such a broad range of alternative explanations, we had to refrain from also controlling mode of data collection. Still, reliance on self-report criteria is a common concern in research on CWB and integrity tests (e.g., Nicol & Paunonen, 2002).

By closer inspection, the degree to which the exclusive use of self-reports may qualify the present results depends on answers to three different questions: (a) To what extent does administering self-report measures of work-related counterproductive behavior affect structural relationships (i.e., differences in correlational patterns between self-report studies and nonself-report studies of CWB)? This question was addressed in a recent meta-analysis by Berry et al. (in press) who report a correlation of .89 between vectors of 18 predictors that were validated against self-reports and nonself-reports of the work-related criterion. (b) To what extent were the effects reported in the present study likely be inflated by the use of self-reports (i.e., common method bias)? This is a serious concern in research on overt integrity tests that include admission scales overlapping with the criterion (Nicol & Paunonen, 2002). However, we did not use any admissions in the present study to measure overt integrity. More applicable to the present study may be findings by Berry et al. (in press), who report that FFM dimensions, on average, correlate no higher with self-report (mean $\rho = .21$) than with nonself-report (.20) criteria of work counterproductive behavior, and Ones et al. (1993, Table 11) who found, for both overt and personality-based integrity tests, that validation studies with similar designs yielded similar results with external and admission criteria of work counterproductivity. (c) To what extent were the present findings possibly *biased* by the use of self-reports (i.e., artificial inflation of the likelihood of supporting our hypotheses, and deflation of the likelihood of rejection)? As our predictions assumed particular advantages of one set of self-report predictors over another set of self-report predictors in some cases and the reverse in others, it seems illogical to expect a systematic bias in favor of our hypotheses due to the use of self-report measures of criteria. In sum, although we are unfortunately not able to rule out empirically the possible impact of method bias on our findings, relevant previous research and logical reasoning do not point to substantial biases.

As a final potential limitation, it needs to be mentioned that our hypotheses were supported in most, but not in all analyses. The only severely contradictory result, however, occurred for the comparison with

the weakest design, which involved relatively small and nonoverlapping samples.

Conclusion and Future Directions

Despite the existence of potential limitations, the comprehensiveness of the present study allows us to draw two major conclusions with some confidence. First, there is now evidence that the relationships of work-related counterproductive behavior with overt and personality-based integrity tests in part stem from different sources. Whereas personality-based integrity tests are well rooted in the FFM, the additional factor of H–H is more important for explaining the validity of overt integrity tests. None of these explanations appears comprehensive, however. Second, we found consistent evidence that findings on integrity tests are not specific to country, language, single instruments, setting, or population. Future research should extend the present results to more established commercial integrity tests (if available), to countries outside the Western cultural sphere, to different criteria (e.g., performance ratings, objective indicators of CWB), and to different operationalizations of personality (e.g., alternative structural models, facet level below FFM or the H–H dimension) and beyond (e.g., attitudes). In light of the present results, we see little reason to expect large differences from our findings with most of these extensions, but this remains a hypothetical statement awaiting empirical tests.

REFERENCES

Alliger GM, Dwight SA. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement, 60*, 59–72.

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

Ashton MC, Lee K. (in press). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*.

Ashton MC, Lee K, Goldberg LR. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology, 87*, 707–721.

Ashton MC, Lee K, Morrison DL, Cordery J, Dunlop P. (2006). An overview of the HEXACO Model of Personality Structure and the HEXACO Personality Inventory. In Lee K, Ashton MC (Chairs), *Revising the five-factor model: a new six-dimensional model of personality*. Symposium at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Ashton MC, Lee K, Perugini M, Szarota P, De Vries RE, Di Blas L, et al. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356–366.

Barrick MR, Mount MK, Judge TA. (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 TE. (1998). Integrity in organizations: Beyond honesty and conscientiousness. *Academy of Management Review, 23*, 154–161.

Berry CM, Ones DS, Sackett PR. (in press). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*.

Brooks P, Arnold DW. (1989). *Reid Report examiner's manual*. Chicago: Reid Psychological Systems.

Cohen J. (1988). *Statistical power analysis for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Costa PT, McCrae RR. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa PT, McCrae RR. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*, 21–50.

Cunningham MR, Wong DT, Barbee AP. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology, 79*, 643–658.

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

Digman JM. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246–1256.

Fortmann K, Leslie C, Cunningham M. (2002). Cross-cultural comparisons of the Reid Integrity Scale in Latin America and South Africa. *International Journal of Selection and Assessment, 10*, 98–108.

Gough HG. (1971). The assessment of wayward impulse by means of the Personnel Reaction Blank. PERSONNEL PSYCHOLOGY, *24*, 669–677.

Hakstian AR, Farrell S, Tweed RG. (2002). The assessment of counterproductive tendencies by means of the California Psychological Inventory. *International Journal of Selection and Assessment, 10*, 58–86.

Hough LM. (1992). The "Big Five" personality variables—construct confusion: Description versus prediction. *Human Performance, 5*, 139–155.

Hu L, Bentler PM. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Hunter JE, Schmidt FL. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

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

Lee K, Ashton MC. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39*, 329–358.

Lee K, Ashton MC. (2006). Further assessment of the HEXACO Personality Inventory: Two new facet scales and an observer report form. *Psychological Assessment, 18*, 182–191.

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

Lee K, Ashton MC, Shin KH. (2005). Personality correlates of workplace anti-social behavior. *Applied Psychology: An International Review, 54*, 81–98.

Lee K, Ogunfowora B, Ashton MC. (2005). Personality traits beyond the Big Five: Are they within the HEXACO space? *Journal of Personality, 73*, 1437–1463.

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. (2003). Attitudes toward personnel selection methods: A partial replication and extension in a German sample. *Applied Psychology: An International Review, 52*, 515–532.

Marcus B. (2006a). *Inventar berufsbezogener Einstellungen und Selbsteinschätzungen [Job-Related attitudes and self-evaluations inventory]*. Göttingen, Germany: Hogrefe.

Marcus B. (2006b). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science, 48*, 226–246.

Marcus B, Funke U, 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.

Marcus B, Höft S, Riediger M. (2006). Integrity tests and the five-factor model of personality: A review and empirical test of two alternative positions. *International Journal of Selection and Assessment, 14*, 113–130.

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

Marcus B, Schuler H, Quell P, Hümpfner G. (2002). Measuring counterproductivity: Development and initial validation of a German self-report questionnaire. *International Journal of Selection and Assessment, 10*, 18–35.

McCrae RR, Costa PT, Piedmont RL. (1993). Folk concepts, natural language, and psychological constructs: The California Psychological Inventory and the five-factor model. *Journal of Personality, 61*, 1–26.

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

Mikulay SM, Goffin RD. (1998). Measuring and predicting counterproductivity in the laboratory setting using integrity and personality testing. *Educational and Psychological Measurement, 58*, 768–790.

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

Murphy KR, Lee SL. (1994b). Personality variables related to integrity test scores: The role of conscientiousness. *Journal of Business and Psychology, 8*, 413–424.

Nicol AAM, Paunonen SV. (2002). Overt honesty measures predicting admissions: An index of validity or reliability? *Psychological Reports, 90*, 105–115.

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

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

Ones DS Viswesvaran C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245–269.

Ones DS, 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 DS, Viswesvaran C, Schmidt FL. (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.

Paunonen SV, Ashton MC. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*, 524–539.

Paunonen SV, Jackson DN. (2000). What is beyond the Big Five? Plenty! *Journal of Personality, 68*, 821–835.

Robinson SL, Bennett RJ. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal, 38*, 555–572.

Rotundo M, Sackett PR. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology, 87*, 66–80.

Ryan AM, Sackett PR. (1987). Pre-employment honesty testing: Fakability, reactions of test taker, and company image. *Journal of Business and Psychology, 1*, 248–256.

Sackett PR, Burris LR, Callahan C. (1989). Integrity testing for personnel selection: An update. PERSONNEL PSYCHOLOGY, *42*, 491–529.

Sackett PR, Wanek JE. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. PERSONNEL PSYCHOLOGY, *49*, 787–829.

Salgado JF. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10*, 117–125.

Salgado JF, Moscoso S, Lado M. (2001, May). *Big Five, integrity and the prediction of self-reported counterproductivity and performance*. Paper presented at the Congress of the European Association of Work and Organizational Psychology, Prague, Czech Republic.

Saucier G, Goldberg LR. (1998). What is beyond the Big Five? *Journal of Personality, 66*, 495–524.

Schmidt FL, Hunter JE. (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.

Schneider RJ, Hough LM, Dunnette MD. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior, 17*, 639–655.

Wanek JE. (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.

Wanek JE, Sackett PR, Ones DS. (2003). Towards and understanding of integrity test similarities and differences: An item-level analysis of seven tests. PERSONNEL PSYCHOLOGY, *56*, 873–894.

Woolley RM, Hakstian AR. (1993). A comparative study of integrity tests: The criterion-related validity of personality-based and overt measures of integrity. *International Journal of Selection and Assessment, 1*, 27–40.