What is integrity and how do we use it? –

Enhancing the validity of integrity by reviewing integrity tests,

expanding the nomological network, and reducing faking

Inaugural-Dissertation

zur Erlangung der Doktorwürde der

Fakultät für Humanwissenschaften

der

Julius-Maximilians-Universität Würzburg

vorgelegt von

Nadine J. B. Schmitt

aus

Würzburg

Erstgutachter:	Professor Dr. Wolfgang Schneider
Zweitgutachter:	Professorin Dr. Tanja Bipp
Tag des Kolloquiums:	30. September 2021

Dedicated to my grandmother Gertrud

Acknowledgements

I would like to express my heartfelt thanks to Professor Dr. Wolfgang Schneider.
I am grateful that he offered me the opportunity to continue my work
on integrity and to finish my doctoral thesis.
Moreover, I am proud to have worked with him and learned from his impressive experience.
Thank you, Professor Dr. Schneider, for all your advice, insights, and continuous support
throughout the years.
I also acknowledge with gratitude Professor Dr. Tanja Bipp for her work as a member of the
dissertation committee and for her help on this thesis.
Her advice and support provided valuable contributions to the dissertation.
Moreover, I acknowledge with thanks
the research contributions of Professor Dr. Anja Göritz,
who shared her thoughts and advice with me.
I thank her for using her data panel to collect some of the data
presented in this dissertation.
I would also like to express special thanks to
Dr. Andrea Beinicke and Dr. Anand Krishna for the extra time
they spent helping me with my dissertation.
Above all, I specially thank my parents, Rosmarie Gagel and Klaus Rindt, for the support
they have given me during my education and work.
There are three very special individuals without whose contributions this dissertation
would not have been possible:
my husband, Simon, and my daughters, Florentine and Valerie.
Simon, thank you for supporting me at every step along the way.
Florentine, thank you for putting a smile on my face while I was working on my dissertation.
Valerie, thank you for giving me time to work by sleeping a lot and playing on your own
so that I could finish my dissertation.

Overview

LIST OF ABBREVIATIONS

AHRD	Academy of Human Resource Development
ARP	Applicant Risk Profiler
ATP	Association of Test Publishers
BIC	Bayesian information criteria
CBI	Counterproductive Behavior Index
COPS	criterion-focused occupational personality scales
CPI	California Psychological Inventory
CWB	counterproductive work behavior
EPPA	Employee Polygraph Protection Act
HEXACO	six-dimensional model of human personality consisting of the traits honesty-humility, emotional stability, extraversion, agreeableness, conscientiousness, and openness to experience
HPI-R	Hogan Personality Inventory–Reliability Scale
IBES	Inventar berufsbezogener und einstellungsbasierter Selbstbeurteilung [Inventory of Job-Related Attitudes and Self-Evaluation]
MMPI	Minnesota Multiphasic Personality Inventory
OCB	organizational citizenship behavior
PRB	Personnel Reaction Blank

SUMMARY

What is integrity and *how do we use it*? These two questions serve as a guide throughout this dissertation. With the aim of answering both questions and thereby contributing to current research on integrity and integrity tests, this dissertation focuses on the construct and criterion validity of integrity tests and aims to enhance both. To accomplish this goal, three approaches were adopted: First, an overview and systematic comparison of integrity tests was conducted with reference to the construction and application of the tests to provide the status quo of integrity tests (Chapter 3). Second, the nomological network of integrity tests was expanded with reference to both personality and behavior at their factor- and facet level (Chapters 4 and 5). Third, two novel methods to reduce faking on integrity tests were tested (Chapter 6).

The overview in Chapter 3 revealed the existence of 76 integrity tests. Scale analysis was conducted on 50 tests, and a systematic comparison was conducted on 16 tests. In line with findings from previous research, the results confirmed that integrity tests are multidimensional and heterogenous. A clear definition of integrity is thus urgently needed.

In terms of the construction of the tests, theoretical models of integrity are missing, as is a clear focus on predicting counterproductive work behavior. Moreover, the tests suffer from overlapping constructs. Regarding the application of the tests, measures include a broad range of criteria aimed at predicting behavior and that take a long time to apply due to the many subscales and many items. Given these key findings, a prototype integrity test could be constructed that would help achieve a clear focus among integrity tests and that would support the interpretation and comparison of findings from different integrity tests within the fields of research and personnel selection.

In Chapters 4 and 5 important correlates of the nomological network of integrity were identified. The personality trait of honesty-humility (r = .45) and its facets of fairness (r = .37) and modesty (r = .34) revealed the most significant relationships to integrity. Moreover, organizational citizenship behavior (r = .37) and its facets of altruism (r = .24), conscientiousness (r = .32), and sportsmanship (r = .36) were found to significantly relate to integrity. Furthermore, integrity tests were able not only to predict organizational citizenship behavior but also to incrementally predict job performance ($\Delta R^2_{\text{factor}}$ = .05; $\Delta R^2_{\text{facets}}$ = .04) and

organizational citizenship behavior ($\Delta R^2_{factor} = .06$; $\Delta R^2_{facets} = .04$) beyond the factor- and facet level of the personality traits of conscientiousness and honesty-humility.

In Chapter 6, two novel methods to prevent socially desirable responding (faking) on integrity tests were investigated: the double-rating method (Hui, 2001) and indirect questioning (Fisher, 1993). While indirect questioning, which only includes the perspective of an other-rating, was not found to prevent faking, the double-rating method, which includes both rating perspectives (i.e., an other-rating and a self-rating) was shown to be able to significantly reduce faking due to two aspects: First, the self-rating of the double-rating method was lower for integrity than was the control group (i.e., self-rating only for integrity without an other-rating) ($F(1, 515) = 3.87$, $p < .05$, partial $eta^2 = .007$), which indicates that less faking on integrity occurs within the double-rating method as compared with a standard self-rating. Second, the difference between the self-rating and the other-rating of integrity within the double-rating method had a significant positive correlation with the impression management scale ($r = .27$, $p < .001$ for the anonymous survey setting and $r = .13$, $p < .01$ for the simulated personnel selection setting), which indicates that participants whose self-rating was more positive than their other-rating for integrity tend to fake their answers to a higher degree.

In conclusion and with reference to the first key question of this dissertation relating to what integrity is, the findings of this dissertation expand our knowledge about integrity, its construct, and its nomological network: Indeed, integrity is a diverse construct that is strongly related not only to the personality trait of honesty-humility but also to organizational citizenship behavior and to the facets of fairness, modesty, altruism, conscientiousness, and sportsmanship. As a result, each of these elements should be considered in the construct of integrity and in its definition.

To answer the second key question of this dissertation, namely how we use integrity, the findings of this dissertation contribute additional knowledge about the predictive validity of integrity tests and about how to protect against faking on these tests. Integrity tests are able to predict organizational citizenship behavior in addition to their previously identified ability to predict counterproductive work behavior and job performance. Moreover, integrity tests can incrementally predict both job performance and organizational citizenship behavior beyond the personality traits of honesty-humility and conscientiousness, two important correlates of integrity. As a result, these predictive results expand the range of criterion validity for integrity tests. Furthermore, the potential applications of integrity tests in

personnel selection can be extended, for example, to include gathering information about the organizational citizenship behavior of future employees. In terms of the problem of faking on integrity tests, the novel method of double-rating, which is simple to apply, may be successful in preventing faking on integrity tests.

In summary, this dissertation contributes new knowledge about the construct and nomological network of integrity, provides a more detailed view on integrity tests and their protection against faking, and expands not only the predictive validity but also the incremental validity of these tests. Main findings of this dissertation aim to provide further insights into research on integrity and integrity tests and into the application of integrity tests in practice.

ZUSAMMENFASSUNG

„Was ist Integrität?“ und „Wie nutzen wir Integrität?“– dies sind die beiden zentralen Fragen dieser Dissertation. Der Fokus liegt dabei sowohl auf der Konstrukt- als auch auf der Kriteriumsvalidität von Integritätstests und deren Erhöhung mit Hilfe von drei Ansätzen: (a) Ein Überblick zu und systematischer Vergleich von Integritätstests mit Bezug auf deren Konstruktion und Anwendung wird erstellt (Kapitel 3). (b) Das nomologische Netzwerk von Integrität wird erweitert mit Bezug auf Persönlichkeit und Verhalten sowohl auf Faktoren- als auch Facettenebene (Kapitel 4 und 5). (c) Zwei neuartige Ansätze, um „Faking“ (Täuschung) in Integritätstests zu reduzieren, werden getestet (Kapitel 6).

In Kapitel 3 wurden im Rahmen einer Literaturübersicht insgesamt 76 existierende Integritätstests gefunden. Für 50 Tests wurde eine Skalenanalyse und für 16 Tests ein systematischer Vergleich durchgeführt. Im Ergebnis zeigte sich, dass die Multidimensionalität und Heterogenität von Integritätstests bestätigt wurde. Integritätstests sind zudem häufig mit anderen Konstrukten konfundiert. Eine klare Abgrenzung und enge Definition von Integrität sind hier dringend notwendig. Im Hinblick auf die Konstruktion der Tests wird deutlich, dass einerseits fundierte theoretische Modelle von Integrität und andererseits ein klarer Fokus auf Vorhersage von beruflicher Kontraproduktivität fehlen.

Bezüglich der Anwendung von Integritätstests wird deutlich, dass diese im Vergleich zur grundlegenden Definition von Integrität einen wesentlich breiteren Kriteriumsbereich bezüglich der Vorhersage von Verhalten haben und—bedingt durch eine Vielzahl von Subskalen und Items—eine große Anwendungsdauer aufweisen. Durch die Zusammenfassung der wesentlichen Ergebnisse der Analysen und Vergleiche von Integritätstests gelingt es, einen Prototyp für Integritätstests zu entwickeln. Dieser Prototyp ermöglicht einen klaren Fokus auf Integritätstests und erleichtert die Interpretation und Vergleichbarkeit von Ergebnissen verschiedener Integritätstests im Anwendungsbereich der Forschung als auch in der Personalauswahl.

In Kapitel 4 und 5 wurden mit Blick auf das nomologische Netzwerk wichtige Korrelate des Konstruktes Integrität erforscht. Die Persönlichkeitseigenschaft Honesty-Humility (r = .45) und deren Facetten Fairness (r = .37) und Bescheidenheit (r = .34) zeigten die höchsten signifikanten Korrelationen mit Integrität. Beim freiwilligen Arbeitsverhalten (r = .37) und deren Facetten Altruismus (r = .24), Gewissenhaftigkeit (r = .32) und Unkompliziertheit (r = .36) wurden ebenfalls signifikante Zusammenhänge

gefunden. Des Weiteren zeigte sich, dass Integritätstests freiwilliges Arbeitsverhalten vorhersagen und inkrementelle Validität besitzen: Integritätstests zeigen inkrementelle Validität über die Faktoren und Facetten von Gewissenhaftigkeit und Honesty-Humility hinaus bei der Vorhersage von beruflicher Leistung ($\Delta R^2_{\text{Faktor}} = .05$; $\Delta R^2_{\text{Facetten}} = .04$) und bei der Vorhersage von freiwilligem Arbeitsverhalten ($\Delta R^2_{\text{Faktor}} = .06$; $\Delta R^2_{\text{Facetten}} = .04$).

Um das Konstrukt Integrität besser vor Konfundierung durch sozial erwünschtes Antwortverhalten (Faking) zu schützen, wurden zwei neuartigen Fragemethoden, die *Doppelte Fragemethode* (Hui, 2001) und *indirektes Fragen* (Fisher, 1993), getestet. Während sich für indirektes Fragen (= Fremdeinschätzung der Items) kein Unterschied im Vergleich zu einer Kontrollgruppe (= Selbsteinschätzung der Items) zeigte, führte die Doppelte Fragemethode (= Fremd- und Selbsteinschätzung der Items) in der anonymen Forschungssituation zu einer signifikanten Reduzierung des Gruppenmittelwertes im Vergleich zur Kontrollgruppe ($F(1, 515) = 3.87$, $p < .05$, partielles $eta^2 = .007$). Dieses Ergebnis legt nahe, dass in der Doppelten Fragemethode weniger Faking im Vergleich zur Kontrollgruppe mit der alleinigen Selbsteinschätzung der Items auftritt. Außerdem konnte gezeigt werden, dass die Differenz in der Antwort zwischen Selbst- und Fremdeinschätzung signifikant positiv mit einem Fragebogen zur Messung von Impression Management zusammenhängt ($r = .27$, $p < .001$ in der anonymen Forschungssituation und $r = .13$, $p < .01$ im simulierten Personalauswahlprozess). Dieses Ergebnis deutet darauf hin, dass Teilnehmer, die eine positivere Selbst- als Fremdeinschätzung hinsichtlich Integrität angaben, dazu tendierten, mehr Faking zu zeigen.

Abschließend und hinsichtlich der Beantwortung der einleitenden Frage „Was ist Integrität?“ bleibt festzuhalten, dass das Wissen über Integrität sowie dessen Konstrukt und nomologischen Netzwerk durch die Ergebnisse dieser Dissertation erweitert werden konnte: Integrität ist ein sehr heterogenes Konstrukt, welches über Persönlichkeitseigenschaften wie Honesty-Humility hinaus auch stark durch Verhaltensvariablen wie freiwilliges Arbeitsverhalten geprägt ist. Integrität ist demzufolge ein heterogenes Konstrukt, welches nicht nur mit der Persönlichkeitseigenschaft Honesty-Humility, sondern auch mit Verhaltensweisen wie freiwilligem Arbeitsverhalten und verschiedenen Facetten wie Fairness, Bescheidenheit, Altruismus, Gewissenhaftigkeit und faires Verhalten verstärkt zusammenhängt. Daher sollten diese Variablen und Facetten sowohl bei einer Definition von Integrität als auch bei Anwendung des Konstruktes in Form von Integritätstests berücksichtigt werden.

Zur Beantwortung der zweiten einleitenden Frage „Wie nutzen wir Integrität?“ tragen die Ergebnisse dieser Dissertation sowohl zur prädiktiven Testvalidität als auch zum Schutz der Integritätstests vor Faking einen Mehrwert bei. Es zeigte sich, dass Integritätstests neben der bisher bekannten Vorhersage von betrieblicher Kontraproduktivität und beruflicher Leistung auch die Vorhersage freiwilligen Arbeitsverhaltens leisten können. Im Weiteren haben Integritätstests den Mehrwert, bei der Vorhersage von beruflicher Leistung sowie von freiwilligem Arbeitsverhalten über Honesty-Humility und Gewissenhaftigkeit als wichtigste Persönlichkeitskorrelate von Integrität hinaus zusätzlich Varianz aufzuklären. Diese Ergebnisse erweitern den Vorhersagebereich der Kriteriumsvalidität von Integritätstests. Die Anwendung von Integritätstests in Bereichen wie der Personalauswahl kann daher verstärkt empfohlen werden, um mehr Aussagekraft über Zielkriterien wie zum Beispiel freiwilliges Arbeitsverhalten zukünftiger Mitarbeiter zu erhalten. Um dabei die Problematik des sozial erwünschten Antwortverhaltens der Probanden zu reduzieren, gibt es erste Hinweise darauf, dass die Doppelte Fragemethode für Integritätstests eine einfach umzusetzende Methode zur Reduzierung sozial erwünschten Antwortverhaltens darstellt.

Abschließend bleibt festzuhalten, dass diese Dissertation neues Wissen über das Konstrukt Integrität und das nomologische Netzwerk von Integrität generiert und einen detaillierteren Blick auf Integritätstests und deren Schutz vor Verfälschung von Antworten durch Faking ermöglicht. Die wesentlichen Ergebnisse dieser Dissertation bringen einen deutlichen Mehrwert für die Forschung zu Integrität und Integritätstests sowie für die Anwendung von Integritätstests in Unternehmen.

CHAPTER 1 INTRODUCTION TO INTEGRITY

One of the most successful entrepreneurs of our times, Warren Buffett, has emphasized the importance of a person's integrity in modern business: "… in looking for people to hire, you look for three qualities: integrity, intelligence, and energy. And if you don't have the first, the other two will kill you…". According to Buffett's statement, integrity is an essential characteristic that every applicant and future employee should have. Otherwise, an employee without integrity tends to steal company equipment, lie about his or her working time, or pass on confidential internal data and thereby harms the organization.

However, what exactly is integrity, and how do we use it? In order to provide a solid basis for answering these two questions, this dissertation begins with an overview of (a) the most important strengths and weaknesses of integrity and of (b) the development of integrity measures. These two overviews are followed by the aims and key questions of this dissertation, which are presented in the third and final part of the chapter.

1.1 STRENGTHS AND DRAWBACKS OF INTEGRITY

Employees who are frequently late or absent, use company resources for personal purposes, or harm colleagues represent a great concern to any organization. This range of counterproductive work behavior (CWB) is a widespread and serious problem in our economic system (Bennett & Robinson, 2000). Such behavior can cost a company large amounts of money, damage its image, and thereby ruin its reputation. These consequences of a lack of integrity can be seen in the myriad scandals of global companies (e.g., Daimler, Siemens) and their top management. For example, the Volkswagen corporation cheated on United States emission tests to improve the apparent performance of its diesel engines. Many companies have faced scandals leading to imprisonment, losses in sales, and penalty fees amounting to billions of euros.

In order to prevent such scandals, employees' integrity must be checked in the setting of personnel selection with a so-called *integrity test*. This kind of test was created in North America more than 70 years ago. Originally, these tests were developed as a screening tool to identify inductees with a criminal background into the Second World War (Ash, 1991;

O'Bannon, Goldinger, & Appleby, 1989). Later, the tests were applied to civil pre-employment screenings to identify applicants who might tend to steal or cheat their employers (Ash, 1991; Marcus, 2000). Based on a broad range of attitudes, personnel characteristics, admissions, and hypothetical behavior, integrity tests can predict counterproductive work behavior in future employees (e.g., Giordano, Ones, & Viswesvaran, 2017).

Integrity tests were quickly developed in the late 1980s and the beginning of the 1990s because the polygraph had been prohibited for use in personnel selection (O'Bannon et al., 1989; U.S. Congress, Office of Technology Assessment, 1983; Camara & Schneider, 1994). Therefore, integrity tests replaced the polygraph as the preferred measure used to identify adequate job applicants (Bergmann, Mundt, & Illgen, 1990). The high demand for and hasty development of integrity tests resulted in their lack of both a definition for their construct and a fundamental theory. Over the years, a debate on the structure of the construct has thereby emerged and remains ongoing (Sackett & Schmitt, 2012; Sackett & Wanek, 1996; Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012b).

Several researchers stated that the construct of integrity is not clearly defined (Sackett & Wanek, 1996; Karren & Zacharias, 2007). Connelly, Lilienfeld, and Schmeelk (2006) explained that "integrity can best be thought of as an open concept, with unclear boundaries and an unclear inner nature." (p. 82). No global operational definition of integrity has yet been provided (Catano, O'Keefe, Francis, & Owens, 2018). T. E. Becker (1998, 2005) theoretically discussed the concept of integrity and mentioned common aspects of integrity tests, which include a tendency to comply with social norms, avoid deviant behavior, and embrace a sense of justice, truthfulness, and fairness.

Barnard, Schurink, and de Beer (2008) described integrity as "a multifaceted and dynamic construct" (p.40). It is the common understanding among researchers that integrity has multiple dimensions (e.g., De Meijer, Born, Van Zielst, & Van der Molen, 2010; Jones, Brasher, & Huff, 2002; Marcus, 2000; Van Iddekinge, Taylor, & Eidson, 2005; Wanek, Sackett & Ones, 2003). Because of this, the construct is difficult to define (De Meijer et al., 2010). Overall, there are five approaches what the construct of integrity might be (Marcus, 2000, 2006):

(a) Integrity can be regarded as a general factor (M. M. Harris & Sackett, 1987; Ones, Viswesvaran & Schmidt, 1993), meaning that it is a construct, such as the g factor of intelligence.

(b) Integrity can be defined as a higher-order construct (Sackett & Wanek, 1996; Catano et al., 2018). In this case, it is referred to as an 'alpha' factor and grouped super ordinately to the personality traits of conscientiousness, agreeableness, and emotional stability. For example, after a compositional correlation, integrity has a score of $r = .97$, which is nearly identical to three factors of the Five-Factor Model (Marcus, 2000).

(c) Integrity can be defined as a composite of facets from the Five-Factor Model (Marcus, Höft, & Riediger, 2006). In detail, for the personality-based part of his integrity test, Marcus et al. (2006) found that two facets of neuroticism (angry hostility, impulsiveness), one facet of extraversion (excitement seeking), three facets of agreeableness (straightforwardness, altruism, compliance), and four facets of conscientiousness (order, dutifulness, self-discipline, deliberation) of the NEO-PI-R (Costa & McCrae, 1992) are relevant. For the overt part of his integrity test, he found that two facets of neuroticism (angry hostility, impulsiveness), one facet of extraversion (excitement seeking), one facet of openness to experience (values), two facets of agreeableness (trust, straightforwardness), and two facets of conscientiousness (dutifulness, deliberation) are relevant (Marcus et al., 2006).

(d) Integrity can be classified as a sixth personality dimension (K. Lee, Ashton & De Vries, 2005). This dimension is termed honesty-humility and lies on the same level as the Big Five. Honesty-humility is described as the quality of being honest and sincere and of having a lack of conceit or greed (Ashton & Lee, 2007). People with high scores avoid manipulating others for personal benefit, are not inclined to break rules, are not interested in wealth and luxuries, and do not strive for a special treatment. In contrast, people with low scores on this scale tend to manipulate others to get what they want, show a tendency to break rules for personal gain, attach importance to material gain, and feel a strong sense of self-importance.

(e) Integrity can be categorized as a compound trait (Hough & Schneider, 1996). Compound traits are measured by empirically chosen facets and basic traits. These components aim to produce maximum prediction for a specific criterion in a specific context, and therefore have a higher criterion-related validity than do basic traits.

To empirically clarify the construct of integrity, structural analyses have been conducted for the items on integrity tests (e.g., Wanek et al., 2003). Different factors have been identified quantitatively (e.g., ranging from 4 to 13 in overt integrity tests), as well as qualitatively (Cunningham & Ash, 1988; Jones & Terris, 1984).

In addition to these test analyses, research has concentrated on determinants of integrity. Regarding the Big Five factors, integrity tests correlate substantially with conscientiousness, agreeableness, and emotional stability (Ones, 1993). With regard to a six-factor model, integrity strongly correlates with honesty-humility (K. Lee, Ashton, & De Vries, 2005). Another personality characteristic that substantially correlates with integrity is self-control (Gottfredson & Hirschi, 1990; Hirschi & Gottfredson, 2000).

Over the years, multiple definitions of integrity have been developed. Nevertheless, none of these definitions has come to be generally accepted for integrity or for integrity tests. Integrity thus remains an imprecisely defined construct that is only known to predict CWB and job performance, though the reasons why this prediction works remain unclear.

Nevertheless, the power to predict CWB and job performance have been discussed within the last decade (e.g., Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012a). Integrity tests have increasingly emerged in personnel selection, which has led to unique challenges for these tests. However, the most challenging issue of integrity remains its underlying construct and the fakability of integrity tests.

Hence, the aim of the present dissertation is to clarify these two challenging issues and in addition to contribute to construct and criterion validity of integrity tests. To achieve this goal, three approaches are used: (a) Integrity tests as the primary measure of integrity are reviewed and systematically compared in terms of their construction and application. (b) The nomological network of integrity is expanded by investigating related as well as predicted variables of integrity. (c) Two novel rating methods are tested to reduce faking, which causes error variance on integrity measures.

To understand the issues regarding the construct of integrity, integrity tests, and their criteria of predictive validity, it is important to consider the historical development of integrity tests. Therefore, an overview is presented in the following section.

1.2 A BRIEF HISTORY OF INTEGRITY TEST DEVELOPMENT

Measuring integrity for personnel selection dates back to 1942. During the Second World War, educational psychologist Gilbert Lee Betts, a personnel consultant of the U.S. Army, developed a screening tool to identify inappropriate inductees with a criminal

background (Ash, 1991; O'Bannon et al., 1989). Betts screened prisoners of the U.S. Army and compared this group with army personnel who had no record of breaking rules or committing illegal actions. He found significant differences in 67 developmental and environmental factors (O'Bannon et al., 1989). Based on these findings, he developed the Biographical Case History. In 1947, Betts recommended his test for personnel selection in business companies. Together with Russel Cassel, he used the test as basis for the Life Experience Inventory (Cassel & Betts, 1956), a measure that constituted the first integrity test designed for civil pre-employment screening (Ash, 1991; Marcus, 2000).

In 1951, lawyer John Reid, who worked for the Chicago Police Scientific Laboratory, published another integrity test based on the polygraph questioning technique (Brooks & Arnold, 1989). The test called Reid Report had originally been developed to support the polygraph (O'Bannon et al., 1989). The items in the Reid Report revealed relations to CWB in regard to attitudes, opinions, and biographical data (Marcus, Funke, & Schuler, 1997). Due to the Reid Report's obvious relationship to CWB, Sackett, Burris, and Callahan referred to this type of tests as an *overt* integrity test (1989).

In addition to the overt measures that refer explicitly to CWB, an indirect integrity measure evolved by removing scales from clinical personality inventories to identify CWB via personality characteristics. For example, scales that identify criminals were taken from the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1951) and from the California Psychological Inventory (CPI; Gough, 1957). The latter inventory formed the foundation for the Personnel Reaction Blank (PRB; Gough, 1971), the first so-called *personality-oriented* integrity test (Sackett et al., 1989). As the relationship between personality characteristics and CWB is not obvious to test takers, personality-oriented tests have also been termed *covert* integrity tests.

Aside from being either an overt or personality-oriented integrity test, a test can be categorized as an integrity test if it meets two criteria (Association of Test Publishers, 2010): First, the test must predict narrow or broad CWB of a person. Regarding this criterion, tests that directly measure CWB by admissions and frequency (e.g., the Counterproductive Work Behavior Checklist from Spector et al., 2006) or tests with an organizational-related instead of an individual-related focus on CWB (e.g., the Corporate Integrity Thermometer from Kaptein, 2008) are not considered integrity tests. Second, the test must have been developed for the purpose of screening applicants or employees. Therefore, integrity-related personality tests with a clinical purpose (e.g., the MMPI from Hathaway & McKinley, 1951)

are not considered integrity tests. Moreover, Ones & Viswesvaran (1998b) added the criterion that integrity tests must be paper-and-pencil, self-reported tests as opposed to tests that utilize other methods (e.g., a polygraph, background check, or interview).

In the late 1980s and at the beginning of the 1990s, a variety of new integrity tests were quickly developed as a result of the Employee Polygraph Protection Act[1] (EPPA) from 27 December 1988 (O'Bannon et al., 1989). The act was based on findings that revealed serious limitations of the polygraph (U.S. Congress, Office of Technology Assessment, 1983). As a consequence, the use of any lie detector test—such as the polygraph and similar physiological measurements (e.g., voice stress analyzer)—to detect applicants' and employees' dishonesty was banned (O'Bannon et al., 1989).

Integrity tests replaced the polygraph as the preferred measure of identifying adequate job applicants, especially in light of validity concerns (Bergmann et al., 1990). In the USA, about 2 million tests were conducted per year in personnel selection, thereby making way for a huge market for integrity tests (Honts, 1991). These integrity tests were therefore hastily developed and were criterion-orientated without using any underlying theory or model and without using a clearly defined construct of an individual's disposition. Originally developed to address a practical need (reducing theft and other counterproductive behavior), integrity tests typically focus more on predicting work behavior than on understanding the construct that the tests measure (e.g., Giordano et al., 2017). These tests thus form a part of criterion-focused occupational personality scales (COPS, Ones & Viswesvaran, 2001).

To broaden the applicability criteria of integrity tests, many studies have focused on different validation criteria of CWB, such as substance abuse (Mastrangelo & Jolton, 2001), absenteeism (Ones, Viswesvaran, & Schmidt, 2003), theft (Ones et al., 1993), and workplace violence (Ones, Viswesvaran, Schmidt, & Reiss, 1994). Moreover, reviews as well as meta-analyses have been compiled with a focus on the validity of integrity tests. One of the most cited studies from this decade is Ones et al.'s meta-analysis (1993). The authors found the validity criterion of CWB to range from ρ = .09 to .46, depending on the particular criterion, the sample, and the type of integrity test. Additionally, researchers identified a prognostic validity of integrity tests for supervisory ratings of overall job performance of ρ = .41 for applicant samples in a predictive design. Despite these impressive findings about

[1] Exceptions are made for security offices such as the FBI (Federal Bureau of Investigation), the NSA (National Security Agency), and the Secret Service, as well as for private security companies or given reasonable suspicion (Schleim, 2008).

validity criteria, it is necessary to expand the predictive power of integrity tests in order to include other types of behavior found in the work environment.

In a meta-analysis, Schmidt and Hunter (1998) found incremental validity in integrity tests for predicting job performance beyond a test of cognitive abilities. This finding is important, especially when it comes to choosing the right combination of methods for personnel selection. Unfortunately, studies on incremental validity caused by integrity tests are rare and nearly non-existent in this research field.

In addition to these positive findings for validity, reviews and meta-analyses have yielded concerns regarding different aspects of integrity tests. Two main concerns have been widely discussed in the literature: (a) the missing definition of the construct of integrity (e.g., Karren & Zacharias, 2007), and (b) the possibility of faking answers (e.g., Alliger & Dwight, 2000; Berry, Sackett, & Wiemann, 2007). In conjunction with these two test aspects, concerns regarding validity issues (i.e. what is measured by integrity tests) have been raised (e.g., Karren & Zacharias, 2007).

Van Iddekinge et al. (2012a) thereby reviewed several validity criteria and—in contrast to Ones et al. (1993)—found reduced validity estimates of integrity tests (Van Iddekinge et al., 2012a): For CWB, the overall mean observed correlation was $r = .26$, in contrast to $r = .33$ as reported by Ones et al. (1993) (when corrected for unreliability in the criterion: $\rho = .32$ vs. $\rho = .47$).[2] For job performance, the overall mean observed validity was $r = .12$, in contrast to $r = .21$ as reported by Ones et al. (1993) (when corrected for unreliability in the criterion: $\rho = .15$ vs. $\rho = .34$).[3] Methodological, structural, and analytical differences between the two meta-analyses have been held responsible for the analyses' discord with regard to validity (W. G. Harris et al., 2012; Ones, Viswesvaran, & Schmidt, 2012; Sackett & Schmitt, 2012). Nevertheless, Van Iddekinge et al.'s meta-analysis (2012a) has initiated a debate about validity as the most important requirement of integrity tests. Furthermore, the heterogeneity of the construct of integrity and respondents' faking on integrity tests have been considered in the debate (W. G. Harris et al., 2012; Sackett & Schmitt, 2012; Van Iddekinge et al., 2012b). Therefore, it is essential to clarify the construct of integrity and develop an effective method to avoid applicant faking.

[2] Both meta-analyses included self-reported and external criteria as well as narrow and broad criteria for validation.

[3] Van Iddekinge et al. (2012a) included standard measures of job performance as criteria of validation, some of which also assessed CWB. The researchers therefore excluded criteria that included CWB. In contrast, Ones et al. (1993) included a variety of validation criteria except for the standard measures (supervisory ratings of overall job performance, production records, and commendations).

To reduce these key concerns, new types of integrity tests continued to be developed, such as the integrity-based conditional reasoning test (Fine & Gottlieb-Litvin, 2013). This test assumes that a person's justification mechanism, which is influenced by an individual's traits, explains his or her behavior (James et al., 2005). The test uses reasoning problems that seem to be solved by choosing a logical reasoning answer. In fact, each answer reflects a justification mechanism (e.g., hostile attribution bias). Moreover, LeBreton, Barksdale, Robin, and James (2007) argued that this kind of test is less prone to faking, and that the construct is not contaminated by other unconscious factors (e.g., self-deceptive enhancement).

Another example of a novel integrity test is the video-based situational judgment test (De Meijer et al., 2010). This test describes a scenario—written or depicted on video—in which a person is caught in a certain kind of conflict. Additionally, some responses that describe an action that can solve the problem are given, and the test taker evaluates the likelihood that the person will decide on a described response. Nevertheless, these novel test formats are not capable of addressing the main problems faced by integrity tests. With regard to answer distortion, faking can harm both criterion-related validity and incremental validity in situational judgement tests (Peeters & Lievens, 2005). In contrast, the use of knowledge instructions (e.g., "What is the correct thing to do?") instead of behavioral tendency instructions (e.g., "What would you do?") could reduce the occurrence of faking (Nguyen, Biderman, & McDaniel, 2005). Moreover, an expansive theoretical basis and a clearly defined construct that underlies situational judgement tests remains missing (Whetzel & McDaniel, 2009).

Keeping in mind the fact that both novel methods have been used for integrity and in the field of personnel selection, more research must be conducted to tackle the problem of construct confusion and the scarcity of theories on integrity measures. Moreover, novel approaches to counteract applicant faking on integrity-based measures must be developed.

1.3 AIMS AND KEY QUESTIONS OF THIS DISSERTATION

Over the years, a great deal of research has been conducted in the investigation into the construct of integrity. In these vastly diverse studies, six main theoretical hypotheses on

the structure of integrity have emerged (Marcus, 2006). In addition to questions on the hierarchical structure of the construct, the dimensionality of integrity has also been investigated multiple times (e.g., Wanek et al., 2003). Nevertheless, while there is agreement on the multidimensionality of the integrity construct, no agreement has yet been reached on the definition or structure of integrity (Karren & Zacharias, 2007).

As a result, researchers have strongly recommended investigating the construct of integrity (e.g., Van Iddekinge et al., 2005; Schmitt & Sackett, 2012) as well as the construct validity of integrity tests (e.g., Sackett & Wanek, 1996). Van Iddekinge et al. stated that "… much more research is needed to increase understanding about what integrity tests measure and whether and how the underlying facets relate to valued criteria." (2012a, p. 520). In addition, Sackett and Schmitt remarked that "future research should address the meaning and nature of the integrity construct for both scientific and applied reasons." (2012, p. 554). It is thereby essential to stabilize the foundation of integrity tests by further investigating their construct validity. Focusing and widening the nomological network of integrity could support this construct validity.

Another helpful approach to stabilizing the foundation of integrity tests could be to focus on the tests systematically. In the past, some overviews of integrity tests have been conducted (i.e., Coyne & Bartram, 2002; O'Bannon et al., 1989; Snyman, Aamodt, Johnson, & Frantzve, 1991), but these overviews are now outdated or have failed to systematically and comprehensively report on integrity tests. For example, in 1998, Ones and Viswesvaran (1998b) reported that 45 commercial integrity tests were available in the United States. It is quite evident that the number of integrity tests has continuously grown since 1998. Therefore, a necessary approach to providing a stable foundation for integrity tests could be to review, analyze, and compare integrity tests systematically and in detail.

In addition to the need to address the construct validity of integrity tests, it is also important to focus on the criterion validity of these tests in terms of work-related behavior. Focusing on the criterion validity has two benefits: (a) Remaining research gaps regarding criterion- and incremental validity can be filled, and (b) the predictive validity of integrity tests can thereby be widened. The predictive validity of integrity tests has been demonstrated multiple times for CWB and job performance (e.g., Van Iddekinge et al., 2012a). In some studies, the predictive validity of integrity tests has also been shown for OCB (e.g., Van Iddekinge et al., 2012a) and even for behavioral integrity ratings (e.g., Marcus, Te Nijenhuis, Cremers, & Heijden-Lek, 2016). Studies on the incremental predictive

validity of integrity tests are rare, but some have yielded positive findings. For example, integrity tests showed to incrementally predict CWB beyond the Big Five traits (Catano et al., 2018) and beyond the HEXACO model (Marcus, Ashton, & Lee, 2013). For predicting job performance, a meta-analysis by Schmidt and Hunter (1998) revealed that integrity tests have the greatest gain in validity over general mental ability tests.

Despite multiple studies on the predictive validity of integrity tests for different work-related behaviors, research gaps in the criterion and incremental validity of integrity tests can be identified: (a) First, in terms of criterion validity, it is unclear whether integrity tests can predict OCB at the facet level. (b) Second, no study that has investigated the gain in validity caused by integrity tests when predicting work-related behavior, took the facet level of predictors and criteria into account. (c) Third, studies on the incremental validity of integrity tests beyond personality traits—which are close to the construct of integrity—are lacking.

From a practical perspective, both kinds of validity (i.e., criterion and incremental validity) have a worthwhile benefit for personnel selection measures when it comes to hiring high-performing employees and, as a result, for contributing to the success of a company (Lievens, Sackett, & Zhang, 2020). Indeed, there is vast empirical support for the relationship between personnel selection and firm performance (Lievens et al., 2020). Therefore, it is essential to fill existing research gaps and increase the power of both the criterion and incremental validity of integrity tests.

Multiple approaches contribute to construct and criterion validity. To provide a stable foundation for researching both kinds of validity, different methods can be combined. Four approaches have been selected for this dissertation with the aim of optimizing the validity of integrity: (a) In the first approach, the nomological network of variables related to the composition of integrity is investigated. (b) In the second approach, integrity tests as a measure of the construct of integrity are further analyzed. (c) In the third approach, an effort is made to reduce the influence of measurement error caused by faking. (d) In the fourth approach, the nomological network of integrity regarding variables that are predicted by integrity is investigated. Each of these four approaches represents a step in the measurement process and is addressed in this dissertation (see Figure 1). The relevant variables and questions for enhancing the validity of integrity are described in the following section.

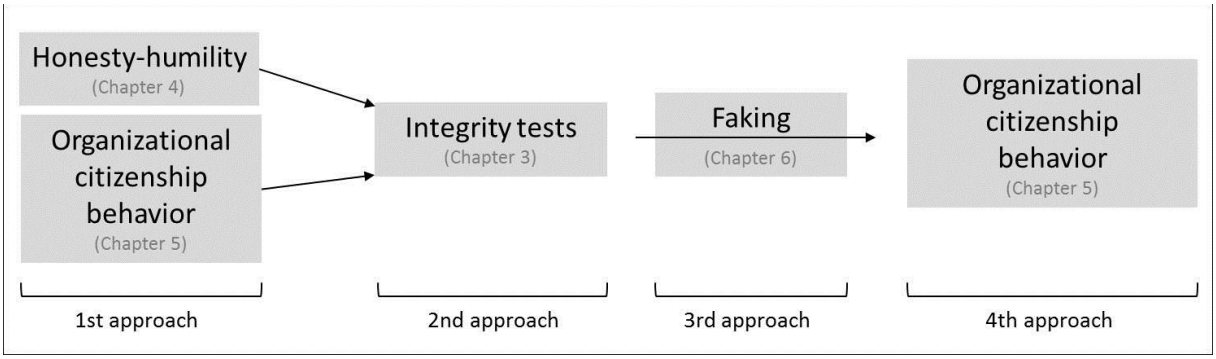

Figure 1. The four approaches to improving construct and criterion validity with reference to the central variables and chapters of this dissertation

The first step of this dissertation begins with a closer look at the construct of integrity via a closer examination of integrity tests (see the second approach). Until now, neither an actual overview nor a detailed and systematic comparison of integrity tests has been established. As a result, the fundamental question of Chapter 3 is: *What is the status quo of integrity tests?* To answer this question, integrity tests were reviewed, subscales were analyzed, and sixteen integrity tests were examined in detail and systematically compared with reference to their construction and application. At the end of Chapter 3, recommendations for key points about integrity tests are made. Based on these key points, a prototypical integrity test is identified and chosen for further research in this dissertation.

The second and third step of this dissertation involve the nomological network of relevant indicators and criteria of integrity (see the first and fourth approaches). In Chapters 4 and 5, relationships between integrity and main constructs are investigated on the factor- and facet level. Moreover, the power of integrity to incrementally predict different work-related behavior is explored on the factor- and facet level. These approaches fill existing research gaps and widen both the criterion and the incremental validity power of integrity tests. To investigate these relationships, a web-based study was conducted that included full- or part-time working employees.

While Chapter 4 focuses on personality traits, honesty-humility—a novel factor that was added to the Big Five—was investigated because the concept of this trait is similar to that of integrity. Because honesty-humility is based on different facets and previous research has failed to focus on integrity and honesty-humility in greater detail, the following two questions arise: *How does integrity relate to facets of honesty-humility, and is the relationship between integrity and honesty-humility influenced by conscientiousness?* As a

result, the relationship between integrity and honesty-humility is investigated on both the factor- and facet level. Moreover, the relationship between integrity and honesty-humility is explored while controlling for one of the most important personality correlates for both traits: conscientiousness. Due to identified research gaps, the following question arises: *Can integrity incrementally predict job performance beyond conscientiousness and honesty-humility on both the factor- and facet level?* All hypotheses presented in this chapter are answered using (partial) correlations and hierarchical regression analyses. In addition, the method of BIC for determining how well a model fits the data in proportion to its number of parameters and the method of relative weight analysis for calculating the relative importance of predictors are integrated.

While Chapter 5 focuses on behavior, organizational citizenship behavior (OCB)—which is strongly related to CWB as the primary criterion of integrity tests—is also considered in the subsequent chapter. OCB is defined as the intent to benefit an organization, support coworkers, and go beyond the formal job requirements of the employment contract (Organ, 1988, 1997). Researchers have called for exploring the relationship between integrity and OCB, but little research has investigated this relationship empirically, and the facet level has been particularly neglected (e.g., Hertel, Bretz, & Moser, 2000). Thus, the following questions arise: *How does integrity relate to facets of OCB,* and *is the relationship between integrity and OCB influenced by conscientiousness?* Due to the structure of OCB, the relationship between integrity and OCB is investigated on both the factor- and facet level. Moreover, whether integrity tests can also predict OCB is examined, which enables the criterion validity of integrity tests to be expanded. Furthermore, the following question arises: *Can integrity incrementally predict OCB beyond honesty-humility and conscientiousness on both the factor- and facet level?* All hypotheses are answered by the same statistical methods as in Chapter 4.

In Chapter 6, this dissertation explored two promising rating methods of effectively measuring the construct of integrity. Due to the nature of integrity measures, integrity tests are vulnerable to applicant faking (Alliger & Dwight, 2000; McFarland & Ryan, 2006), which is still an unsolved issue of integrity tests (Karren & Zacharias, 2007).

Previous studies have revealed that faking can diminish the validity of measurements (e.g., Griffith & Peterson, 2008; Tett & Christiansen, 2007), significantly muddle the rank order of participants (e.g., Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Griffith, Chmielowski, & Yoshita, 2007; Peterson, Griffith, & Converse, 2009) and–in the worst-case

scenario–influence the choice of suitable applicants (e.g., Morgeson, 2004). Due to these research findings and to the wide range of existing methods for combatting faking, Burns and Christiansen (2011) stated that to date, no method of avoiding applicant faking has proven sufficient. Moreover, researchers have recommended that future research should focus on methods to reduce faking on integrity tests (Karren & Zacharias, 2007).

The double-rating method has yielded promising results for reducing faking in measurements of socially desirable responding (Hui, 2001; Thomas, Grawitch, & Scandell, 2007). The method of indirect questioning has also revealed promising for reducing faking on CWB measurements (Dalal & Hakel, 2016). Hence, these two methods are adapted to integrity testing, and the following central question arises in Chapter 6: *Can the double-rating method and indirect questioning effectively prevent faking on integrity tests?* Moreover, as the underlying mechanism of both methods has not been thoroughly investigated, the following question arises: *Do the double-rating method and indirect questioning yield different levels of integrity test scores when respondents refer to the reference group of the majority in contrast to the reference group of the minority?*

To answer these questions, a web-based study was conducted by a university-based online panel from Germany. A dataset of 1,450 people with an average age of 44 years (*SD* = 10.70) was involved in the study. The first method, which was used to reduce faking on integrity tests, was the double-rating method by Hui (2001). In the original study, Hui (2001) asked participants in the first step to answer the question, “Which answers do you think people would choose in order to appease others?” After completing the full questionnaire, in the second step, participants answered the question, “How accurately do the statements describe your own behavior and attitudes?” With reference to the novel application of this method, participants answered integrity-related items while considering the possible answer of others regarding the guiding question of “what might another (/other) person (people) answer?” In addition, participants filled out the same integrity-related items while considering only their own answers. The second method that was used to prevent faking was indirect questioning (Fisher, 1993). This method only involves the rating perspective of others via judging the behavior of others (“How would others act?”); this perspective is contrary to the standard rating of tests from one’s own perspective (“How would you act?”).

In order to systematically investigate both methods, an experimental study with a between-subjects design was used. In the study, the method (indirect questioning vs. double-rating method), the reference group (other rating with reference to a single person

vs. reference to a group of people), and the situation (anonymous survey vs. simulated personnel selection situation) were varied. An additional control group was used in the form of a self-rating only. All hypotheses are analyzed using the statistical method of variance analysis.

In summary, all key findings from Chapters 3 to 6 are discussed with a focus on the validity of integrity. Main findings relating to the two central questions of this dissertation—namely "What is integrity?" and "How do we use integrity?"—and the key questions introduced in this chapter are discussed (see Figure 2). Furthermore, implications are made for research as well as practice. Strengths and limitations of the studies are additionally analyzed in order to propose implications for future research that supports a clear construct of integrity.

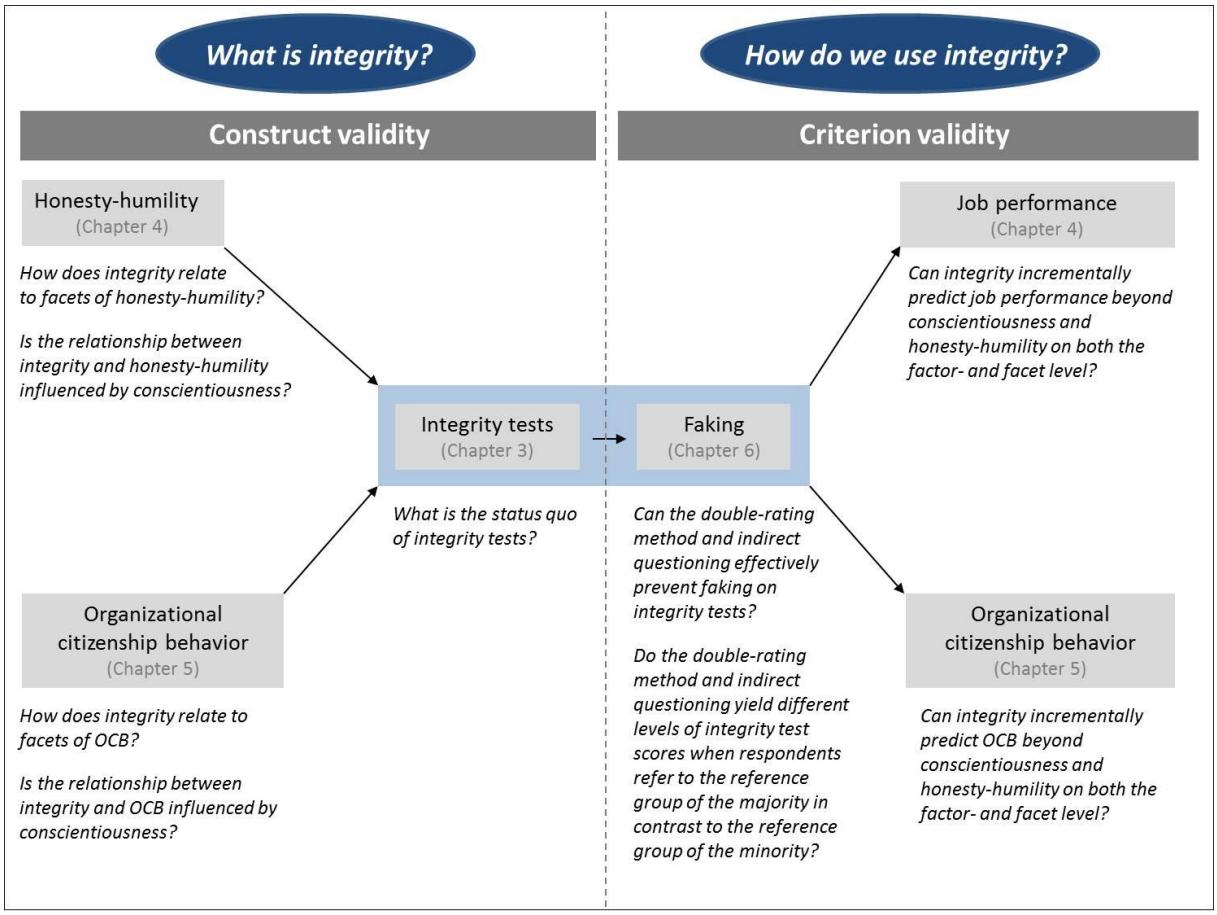

Figure 2. Overview of the focus of this dissertation, including its two key questions on construct and criterion validity, and its key questions with reference to the central variables of this dissertation

CHAPTER 2 THEORETICAL FRAMEWORK OF INTEGRITY AND INTEGRITY TESTS

In this chapter, the theoretical framework of integrity is discussed. In the first part of the chapter, theoretical and empirical definitions and concepts of integrity and its structure are summarized. In addition, data about the construct validity of integrity are presented. Based on these definitions, theoretical approaches of integrity are presented in the second part of the chapter. In the third part, empirical data on the determinants of integrity that are part of the nomological network of integrity are presented. In the final part of this chapter, the definition of CWB, different approaches to its structure, and its determinants as the primary criterion of integrity tests are described.

2.1 THE CONSTRUCT OF INTEGRITY

Integrity is a theoretical construct that cannot be observed, but rather inferred by different observable indicators. Based on these indicators, the concept of integrity has been frequently described by researchers.

At the beginning of the investigation in integrity, researchers often used the term integrity as a synonym for honesty (e.g., Sackett & Harris, 1984) because tests often focused on predicting theft at work for personnel selection (Marcus, 2000). Although honesty is a multidimensional construct (involving e.g., the aspects of cheating, lying, stealing), integrity is a broader concept that includes additional aspects, such as dependability and compliance with norms. Therefore, Sackett et al. (1989) introduced the term *integrity*.

Some researchers initially referred to integrity as a moral concept and a personality trait that describes the accordance of behavior and moral standards (e.g., G. K. Becker, 2009; McFall, 1987), whereas others referred to integrity as a behavioral concept that displays the congruency of words and actions (independently of a society's assessment of a person's values) (e.g., Simons, 2002).

The Academy of Human Resource Development emphasized integrity as one of the values relevant to the field of human resources. In addition, the institution helps to explain integrity via a list of characteristics: honesty, fairness, respect for others, awareness of personal values, belief systems, needs and limitations, clarification of values, and avoidance

of potential conflicting relationships (Academy of Human Resource Development [AHRD], 1999).

Over the years, integrity has been used with different meanings, perspectives, and scopes. As a clear definition of integrity has not yet to be agreed on, theoretical concepts and empirical data on integrity are further described below.

Concepts of integrity

In the following section, an attempt is made to highlight the most critical underlying concepts of integrity (see Table 1). The term integrity has been defined in different fields of science (e.g., criminology) as well as in different sub-fields (e.g., business crime). Many authors have described integrity via a moral or ethical approach.

In 1998, T. E. Becker identified integrity as a principal determinant of trust in organizations. With regard to objectivism, he described integrity as a behavior that is in accordance with a morally justifiable value system and additionally includes components such as sociability, thrill-seeking, and conformity (Nicol, 1999). T. E. Becker (1998) accordingly focused on integrity as a person's commitment to universal principles and viewed integrity as consisting of trustworthiness as well as the congruence of behavior and words.

In contrast to this intrapersonal perspective of integrity, Trevinyo-Rodríguez (2007) defined integrity as an organized multileveled framework that integrates personal integrity, moral integrity, and organizational integrity. However, the concept of integrity in the sense of integrity tests that identify people might be involved in CWB is not related to moral development or structure (Marcus, 2000).

Regarding the prediction of CWB, Murphy (2000) described integrity as a broader concept "including a willingness to comply with rules, internalized values, norms, and expectations." (p. 266). This definition indicates a focus on both internal (i.e., willingness, values) and external factors (i.e., rules and norms).

Barnard et al. (2008) developed a well-grounded, complex concept of integrity. Based on ten in-depth interviews with South African business leaders, Barnard et al. (2008) created a conceptual framework of integrity. They described integrity as being "conceptualised as a multifaceted and dynamic construct based on a moral foundation and inner drive that is managed by cognitive and affective processes manifesting various integrity-related

behaviours." (p. 40). Integrity thus comprises five interacting components: (a) the moral compass (fundamental values, principles, norms), (b) the inner drive (personal motives and ideals), (c) (intra- and interpersonal) authenticity (in reference to the moral compass and the inner drive), (d) cognitive functions (moral reasoning), and (e) affective functions (moral feeling) (Barnard et al., 2008). In addition, researchers explained that integrity is influenced by certain developmental contexts, such as parental role models or religion. Moreover, eight competencies have been identified that reflect behavioral manifestations of integrity: self-motivation and drive, moral courage and assertiveness, honesty, consistency, commitment, diligence, self-discipline, responsibility, trustworthiness, and fairness.

Table 1

Summary of integrity concepts

Author(s)	**Description of integrity**
Barnard, Schurink, & De Beer (2008)	• multifaceted and dynamic construct based on a moral compass and inner drive • managed by cognitive and affective processes that manifest various integrity-related behaviors
T. E. Becker (1998)	• congruency of behavior and a morally justifiable value system
Mayer, Davis, & Schoorman (1995)	• a person's commitment to universal principles • integrity as a key component of trustworthiness and congruence of behavior and words
Murphy (2000)	willingness to comply with rules, internalized values, norms, and expectations
Noelliste (2013)	• personal integrity • moral integrity • organizational integrity
Trevinyo-Rodríguez (2007)	integrity as a trait in a moral (ethical) context

Note. Alphabetical order by surname of author(s).

Noelliste (2013) described integrity as consisting of three components: (a) personal integrity, which is a person's own congruence (e.g., a person's words and behavior), (b) moral integrity, which is congruence with a society's principles, and (c) professional integrity, which is congruence with the principles of a company (e.g., the code of conduct).

Empirically based descriptions of integrity

There are quite a number of approaches to defining what integrity might be. In addition to a framework of theoretical descriptions and concepts, some hypotheses are supported by empirical data. In sum, there are five empirically driven descriptions of integrity identified in the literature. These descriptions are presented and discussed in the following section (see Table 2):

Table 2

Overview of empirically based definitions of integrity

Hypothesis	*r*	Studies
Integrity is conscientiousness.	.36 .42	Murphy & Lee (1994a) Ones et al. (1993)
Integrity is a facet of conscientiousness named self-control.		Sackett & Wanek (1996)
Integrity is a *g* factor of personality that comprise of conscientiousness, emotional stability, and agreeableness.	.33-.42 .28-.31	Ones (1993) Marcus et al. (1997)
Integrity is a "compound trait" of several facets of basic personality traits.		Hough & Schneider (1996) Marcus et al. (1997)
Integrity is a sixth factor of personality (named honesty-humility).	.50 .66	K. Lee, Ashton, & De Vries (2005) Marcus, Lee, & Ashton (2007)

Note. Based on *Kontraproduktives Verhalten im Betrieb: Eine individuumsbezogene Perspektive [Counterproductive behavior in organizations: An interindividual perspective], by* B. Marcus, 2000, Göttingen: Verlag für Angewandte Psychologie. Copyright 2000 by Hogrefe Verlag.

With regard to the Big Five personality dimensions, conscientiousness displays the strongest relationship with integrity (e.g., Ones et al., 1993). As a result of this finding, some researchers place integrity on a par with conscientiousness. However, further research indicated that partialling conscientiousness out of integrity only had a small effect on the magnitude of the correlation of integrity with job performance, whereas controlling for integrity considerably reduced the correlation between conscientiousness and job performance (Murphy & Lee, 1994a; Ones, 1993). Thus, although there is a strong empirical

relationship between integrity and conscientiousness, integrity is also influenced by additional, related facets.

Taking a closer look at the facet level of conscientiousness, the facet of self-control is important in the relationship between both constructs. While Big Five measures emphasize other facets of conscientiousness (e.g., orderliness, perseverance, and conformity, Hogan & Ones, 1997), integrity is strongly related to self-control (Sackett & Wanek, 1996). Thus, Sackett and Wanek (1996) hypothesize that self-control is responsible for the incremental validity of integrity tests over measures of conscientiousness. Despite this strong relationship, conscientiousness and its facets fail to account for all the variance in integrity.

Regarding high correlations of integrity with personality traits, researchers suggest that integrity is a conglomerate of the three Big Five traits of conscientiousness, agreeableness, and emotional stability. Each of these three traits consistently demonstrates a moderate correlation with integrity that is stronger for personality-based than for overt integrity tests (Marcus et al., 1997; Ones, 1993). Moreover, composite correlation reveals a maximum of .97, when specific variance is eliminated.

Researchers also categorize integrity as a ”compound trait” based on its high correlations with facets of conscientiousness, agreeableness, and emotional stability (Sackett & Wanek, 1996; Salgado, Moscoso, & Berges, 2013). The concept of such a compound trait is based on empirically related facets of traits that are consistently identified in studies and offer a maximum criterion-related validity (Berry, Sackett, & Wiemann, 2007).

Beyond the five factors of personality and their facets, a sixth factor has been developed: honesty-humility (Ashton & Lee, 2001). This trait refers to being honest and genuine as well as showing fairness and cooperative behavior (Ashton & Lee, 2007). Studies revealed that honesty-humility is substantially related to integrity (K. Lee, Ashton, & De Vries, 2005; Marcus, Lee, & Ashton, 2007). K. Lee, Ashton, and De Vries (2005) highlighted honesty-humility being remarkable for integrity tests based on studies using overt rather than personality-based integrity tests. The relationship between honesty-humility and overt integrity test scores tends to be stronger than that between honesty-humility and the Big Five traits. Therefore, conceptionally and empirically, integrity is clearly linked with honesty-humility.

With this criterion-focused approach, integrity test construction focuses on variables that predict CWB as the primary criterion. Integrity thereby refers to a multitude of

variables, such as attitudes, values, behaviors, and admissions, which are reflected in heterogeneous subscales of integrity tests. As a consequence, analyses of integrity test items have been conducted. The identified components of the tests provide support for the idea that integrity tests are multifaceted and that there may be a hierarchical construct underlying these tests can be suggested (Berry, Sackett, & Wiemann, 2007).

Dimensionality of integrity tests

The scales of overt as well as personality-based integrity tests reveal a broad range of themes. To underline the multidimensional character of integrity, factorial analyses have been conducted with a large number of particularly overt integrity tests. However, factor analyses revealed different solutions for overt and personality-based integrity tests (see Table 3).

Regarding overt integrity tests, Cunningham and Ash (1988) identified four factors for the Reid Report. M. M. Harris and Sackett (1987) also identified four major components for the Personnel Selection Inventory, whereas Jones and Terris (1984) found six components for the same overt measure. Moreover, W. G. Harris (1987) found six factors for the Stanton Survey.

Table 3

Overview of dimensionality research for integrity tests

Author(s)	Integrity test(s)	No. of factors	Themes of factors
Cunningham & Ash (1988)	Reid Report	4	1) self-punitiveness 2) punitiveness towards others 3) self-projection 4) projection towards others
M. M. Harris & Sackett (1987)	PSI	4	1) temptation and thoughts about dishonest behavior 2) actual or expected dishonest activities 3) norms about the dishonest behavior of others 4) impulse control and reliable tendencies

W. G. Harris (1987)	Stanton Survey	6	1) general theft 2) opportunism 3) employee theft 4) leniency 5) perceived pervasiveness of dishonesty 6) association with dishonest individuals
J. Hogan & Brinkmeyer (1997)	Reid Report, ERI, HPI	4	1) punitive attitudes 2) illegal drug use 3) reliability 4) theft admissions
Jones & Terris (1984)	PSI	6	1) theft temptation and rumination 2) theft rationalization 3) projection of theft in others 4) theft punitiveness 5) inter-thief loyalty 6) personal theft admissions
O'Bannon, Goldinger, & Appleby (1989)	overall	4	1) admissions of illegal or disapproved activities 2) opinions towards illegal or disapproved behavior 3) descriptions of one's own personality and thought patterns 4) reactions to hypothetical situations
Wanek, Sackett, & Ones (2003)	PSI, Reid Report, Stanton Survey, ERI, PRB, PDI-EI, IPI	4	1) antisocial behavior 2) socialization 3) positive outlook 4) orderliness/ diligence
Woolley & Hakstian (1992, 1993)	Reid Report PRB, PDI-EI, ERI	4	1) conventional commitment 2) intolerance for dishonesty 3) socialized control 4) active conscientiousness

Note. Alphabetical order by author(s). Overt integrity tests: PSI = Personnel Selection Inventory, Reid Report, Stanton Survey. Personality-based integrity tests: ERI = Employee Reliability Index, HPI = Hogan Personality Inventory, IPI = Inwald Personality Inventory, PRB = Personnel Reaction Blank. Mixed (overt and personality-based) integrity test: PDI-EI = Personnel Decisions Incorporated, Employment Inventory.

With a focus on multiple overt as well as personality-based integrity tests, several researchers found a quantitatively homogeneous (four factors) but qualitatively heterogeneous solution (e.g., O'Bannon et al., 1989; Woolley & Hakstian, 1992, 1993). J. Hogan and Brinkmeyer (1997) also identified four factors for three integrity tests via principal component analysis, but revealed another (latent) factor *conscientiousness* via

confirmatory factor analysis. Wanek et al. (2003) investigated seven integrity tests and found 23 different themes, which were clustered into four major components of integrity tests via a principal component analysis. Due to the results found when using principal component analysis, researchers suggested that integrity tests are characterized by multidimensionality.

However, although many studies have found a four-factor solution, the labelling of the components differs. Categories that are rife with especially overt integrity tests are (a) the observed distribution of dishonest behavior, (b) rationalizations of and excuses for CWB, (c) hypothetical thinking about CWB, (d) punitiveness and tolerance of CWB in others, and (e) beliefs about widespread CWB in the general public (including judging dishonest behavior as acceptable) (Marcus, 2000, 2006). Moreover, the category of impulse control was supplemented by Murphy (2000) in his summary of common dimensions of integrity tests.

To strengthen the assumption of multidimensionality, correlations of different integrity tests are taken into consideration. Regarding the convergent validity of integrity tests, overt integrity tests have a mean correlation of $\rho = .45$, whereas personality-based integrity tests have a mean correlation of $\rho = .70$ (Ones et al., 1993). This finding indicates that overt integrity tests refer to a greater range of themes than do personality-based integrity tests. In addition, one meta-analytic finding revealed a mean correlation between overt and personality-based integrity tests of $\rho = .39$ (Ones et al., 1993). A later study, which included seven integrity tests, found a correlation between both kinds of integrity tests with an average of $r = .52$ (Sackett & Wanek, 1996). This finding indicates that the two kinds of tests do not measure the same construct.

In summary

Regarding concepts of integrity, empirically based descriptions and studies investigating the dimensionality of this construct, several conclusions can be reached. There are references indicating that integrity is a multidimensional construct (e.g., J. Hogan & Brinkmeyer, 1997; Wanek et al., 2003). While integrity can refer to both a higher-order construct and a compound trait consisting of personality facets, its structure seems to lie on different hierarchical levels (Berry, Sackett, & Wiemann, 2007). Moreover, several studies have demonstrated that integrity tests have construct validity (e.g., Ones et al., 1993; Mumford, Connelly, Helton, Strange & Osburn, 2001; Wanek et al., 2003). Thus, despite the

lack of theoretically guided test development, empirical evidence for the sufficient construct validity of integrity tests is available.

While a widely accepted definition of integrity remains missing (Karren & Zacharias, 2007), it is critical to define integrity when conducting research on the construct. A definition of integrity is needed for this dissertation for three reasons: First, a generally accepted definition of integrity is not yet available. Second, the definition of integrity provides a content framework for the research conducted in this dissertation. Third, the definition of integrity helps to interpret the results of this dissertation and to generalize its findings. As a result, integrity is defined for this dissertation as follows:

Integrity is a compound construct based on key personality traits (i.e., conscientiousness, self-control), attitudes (i.e., fairness), and work-related behavior (i.e., supporting colleagues). It is defined to follow laws, organizational rules, and own individual rules, which are based on individual attitudes of what is right and what is wrong. Therefore, one of integrity's key components is compliance with rules. Moreover, integrity is clearly related to work and organizations. In addition, individual attitudes and behavior involves acting correctly toward others as well as to not insisting on advantage for oneself.

2.2 THEORETICAL APPROACHES OF INTEGRITY

In general, integrity tests have no underlying theory in their construct. In some tests, there is a theoretical fragment called the *false consensus effect*. Moreover, there are many theories from various areas of expertise (e.g., social psychology, criminology) that relate to familiar constructs of integrity and have influenced the development of integrity tests (e.g., self-control).

False consensus effect

The false consensus effect is a bias of social perception and attribution process originating from social psychology, in which a person estimates a higher frequency for an own characteristic (Ross, Greene, & House, 1977). The characteristic can refer not only to a behavior or personality trait but also to an attitude, opinion, judgement, or feeling. For example, an honest person will overestimate the ratio of honest people in public while

judging dishonesty to be more uncommon. In contrast, a dishonest person will estimate dishonesty to be more common in public and underrate the proportion of honest people.

The false consensus effect has been empirically confirmed by an impressive meta-analysis that revealed a high statistical significance and a large effect size (Mullen et al., 1985). In addition, Mullen et al. (1985) investigated methodological, situational, and procedural variables. They found that the effect is influenced by the sequence of answering as well as the number of questions: Both estimating before making a behavioral choice and answering fewer items leads to more significant results in favor of the false consensus effect. However, the effect is independent of the degree of the generality of the reference group as well as of the actual difference in the consensus. Neither variables significantly predicted study outcomes. Mullen et al. (1985) argued that these results supporting the false consensus effect are not self-presentational or motivationally driven.

In general, there are two viewpoints regarding the function of the false consensus effect (Ross et al., 1977). Some researchers assume that the bias has the function of self-defending or reducing dissonance in the context of attribution processes. However, other researchers believe that the effect has no motivational or strategic basis or result from a selective exposure effect because people tend to approach other with same interests and attitudes. We thus more strongly recall people who have same personality features that we do. Nevertheless, the underlying mechanism of the false consensus effect has not yet been clarified.

Independently of the fundamental basis of this bias, the effect provides a valuable assessment of a person's integrity. People who act counterproductively at work consider CWB to be widespread because this behavior is common for them. The false consensus effect has thereby become a part of many integrity tests. As a further example, a typical item of these scales is "Nearly everyone has stolen something when the opportunity was favorable". These items refer not only to working situations but also to situations in everyday life (e.g., "Nearly everyone cheats on the annual tax declaration."). A person who scores high on these items is categorized as having low integrity, whereas a person who scores low on these items is classified as having high integrity. Finally, the false consensus effect is frequently used by scales in overt integrity tests, e.g., the *Inventar berufsbezogener und einstellungsbasierter Selbstbeurteilung* [the Inventory of Job-Related Attitudes and Self-Assessments] by Marcus (2006).

Theory of planned behavior

The theory of planned behavior (Ajzen, 1985, 1991) aims to predict human behavior in specific situations. It was derived from the theory of reasoned action and developed further by adding perceived behavioral control as an additional determinant of intentions and behavior (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975). The theory is thus common and aims to explain behavior that occurs without behavioral control (see Figure 3).

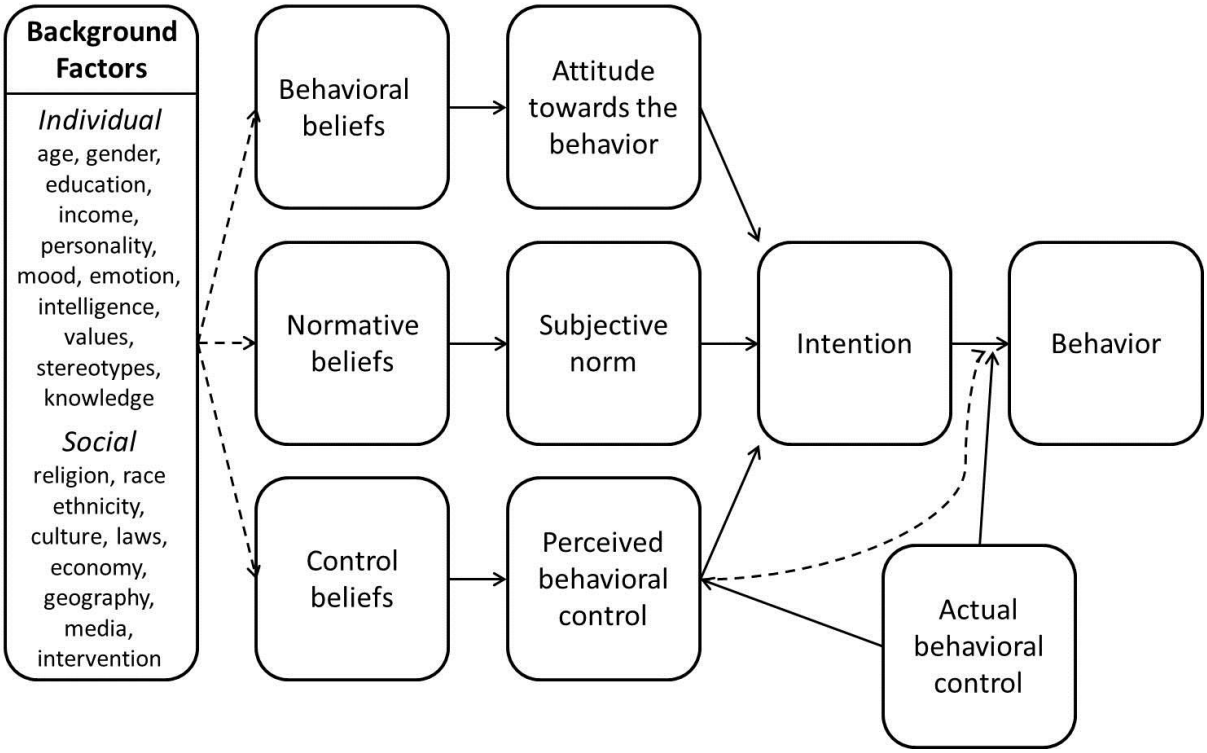

Figure 3. The theory of planned behavior. From "Theory of planned behavior with background factors", by I. Ajzen, 2017 (https://people.umass.edu/aizen/tpb.background.html). Copyright 2019 by Icek Ajzen.

Both personality traits and situational variables play a role in predicting behavior. The theory of planned behavior indicates that behavior is best predicted by behavioral intention, which is influenced by three components: (a) attitudes toward the behavior, (b) the subjective norm, and (c) perceived behavioral control. Moreover, these three categories of beliefs are influenced by different background factors, such as individual factors (e.g., personality traits), and social factors (e.g., religion) (Ajzen, 2017). Their impact on specific behavior is thereby more indirectly influenced.

The theory of planned behavior is empirically strengthened in many areas of psychology (Ajzen, 1991; Ajzen & Madden, 1986; Reinecke, Schmidt, & Ajzen, 1996) and has even been used to predict theft (Beck & Ajzen, 1991). With regard to the behavior of integrity as well as counterproductive behavior, the theory describes the relevance of situational factors: First, the attitudes toward the behavior of integrity and CWB are relevant. This is in line with overt integrity tests, which focus on counterproductive attitudes, such as the punitiveness of others or hypothetical thinking about CWB. Second, the norm of the individual is crucial to his or her behavior regarding integrity and CWB. Research indicates that people who display CWB have a different system of norms than people with high integrity (Marcus, 2006). Third, perceived behavioral control, which means being confident regarding own's ability to perform, indicates that people consider how successful their action might be.

Regarding intentions of behavior as well as perceived behavioral control, which are two major components of acting, the theory of planned behavior is more closely aligned with overt integrity tests than with personality-based integrity tests. Moreover, test authors refer to the theory of planned behavior (Jones & Terris, 1991; referring to Marcus et al., 1997). While the focus of the theory relies on situational circumstances that account for counterproductive behavior, other theories emphasize personality traits to explain the occurrence of counterproductive behavior.

The theory of self-control

The theory of self-control was developed by Gottfredson and Hirschi (1990) as a general theory of crime in the field of criminology. Based on an extensive literature review, Gottfredson and Hirschi recognized that all types of crime are positively related over a long time period. The authors thus searched for a trait that could explain their empirical findings by analyzing common features of criminal behaviors. They were thereby able to identify a lack of self-control as the underlying key core of crimes, which led to the development of their theory of self-control.

This theory states that individuals with low self-control do not consider the long-term consequences of their behavior. As a result, individuals cannot wait for gratification and act impulsively. For example, they drop out of with school and thereby miss the chance to find a good job with a good salary. In contrast to people with low self-control, individuals with high

self-control avoid actions that provide an advantage in the short term, but high costs in the long term. High self-control is thus the ability to focus on the long-term costs of behaviors (Gottfredson & Hirschi, 1990) and is negatively associated with facets such as impulsivity, callousness, and risk-taking tendencies (Bock, 2007).

The theory has been empirically confirmed, and the characteristic of self-control is associated with a scope of different counterproductive behaviors (Hirschi & Gottfredson, 2000). With regard to work behavior, Marcus and Schuler (2004) found a substantial negative relationship between self-control and CWB (r = -.63). In addition, the authors discovered an incremental validity of self-control beyond age and situational variables for predicting CWB (ΔR^2 = .365). As Marcus and Schuler stated: "Thus, there is clear evidence from hierarchical regressions that self-control was by far the best predictor of GCB [general counterproductive behavior] among all variables under research in this study." (2004, p. 654). Moreover, the finding that self-control is a highly significant predictor of organizational and interpersonal deviance was confirmed (e.g., Bechtold, Welk, Harting, & Zapf, 2007).

While a lack of self-control is associated with CWB, high self-control can be related to integrity. A positive and significant correlation of r = .43 was found between self-control and integrity (Bazzy, Woehr, & Borns, 2017). This finding was supported by Marcus and Schuler (2004), who found significant relationships between self-control and overt as well as personality-based subscales of an integrity test. Moreover, self-control even added validity beyond conscientiousness and the other Big Five personality dimensions (Bazzy et al., 2017). Researchers therefore concluded that self-control belongs to the nomological network of integrity and is "a key component for understanding integrity" (p. 102, Bazzy et al., 2017).

2.3 INTEGRITY TESTS PREDICTING COUNTERPRODUCTIVE WORK BEHAVIOR

Integrity tests aim at identifying people who tend to exhibit CWB. This definition indicates a criterion-orientated focus on these measures. Moreover, integrity tests obviously predict a vast range of CWB criteria. Therefore, it is necessary to conduct a detailed examination of CWB, including its definitions and categories as well as its most important theories and correlative variables.

Definition of CWB

There are many definitions for the construct of CWB (Nerdinger, Blickle & Schaper, 2008). In contrast to a missing consistent definition for integrity, all definitions of CWB includes three main issues (Nerdinger, 2008; Marcus & Schuler, 2004):

(a) CWB damages the legitimate interests of an organization. This means that CWB is of concern to the law. For example, if a company acts illegally and an employee fights against this illegal action, this would not be a case of CWB. (b) CWB is intentional behavior. This means that it excludes unintentional behavior. For example, if an employee unintentionally damages a company's property, this would not be a case of CWB. (c) CWB includes an intention to harm members of the organization or the organization itself. This means that CWB focuses on the behavior itself rather than on the damage or consequences caused by the behavior. For example, if an employee steals secret information from a company and the company immediately detects the theft without any costs or disadvantages, this would represent a case of CWB.

Based on this definition, CWB needs to be clearly separated from other concepts. For example, delinquency refers to actions that are connected with legislation and therefore do not capture all kinds of CWB (Lamnek, 1993, based on Marcus, 2000). In addition, antisocial behavior does not include all kinds of CWB. For example, providing internal data to a friend would not be captured by the concept of antisocial behavior because this behavior is widely socially accepted. These mentioned concepts comprise a narrower range of actions, whereas the construct of CWB is multidimensional due to its different forms, which can include a variety of actions, deviant people, and victims (Gruys & Sackett, 2003).

Structure of CWB

In the consideration of the variety of CWB, different approaches to structuring the construct of CWB can be taken (Marcus, Taylor, Hastings, Sturm, & Weigelt, 2016). The four most well-known and major approaches are introduced below:

First, an important classification for CWB was introduced by Hollinger and Clark (1982, 1983), who distinguished between two forms of CWB: *property deviance* and *production deviance*. The first form refers to actions against organizational or employees' assets (e.g., theft), and the latter refers to harm caused by organizational norms related to work (e.g., absenteeism). Behavior against an organization is thus included in both forms of

CWB. Despite this two-categorical framework of CWB, not all kinds of CWB are included (e.g., aggression against colleagues is excluded).

Second, Robinson and Bennett (1995) introduced a two-dimensional solution based on rating and categorizing similarity between pairs of behavior to multidimensional scaling (see Figure 4). The first dimension refers to the target of CWB, which differs in the organization or in (a) member(s) of the organization (interpersonal). The second dimension refers to the severity of the action as a continuum from minor to serious. Based on these two dimensions, the four quadrants of CWB are (a) *property deviance* (serious behavior against the organization), (b) *production deviance* (minor behavior against the organization), (c) *personal aggression* (serious interpersonal behavior), and (d) *political deviance* (minor interpersonal behavior). By developing a measure for CWB, Bennett and Robinson (2000) demonstrated the construct validity of their classification.

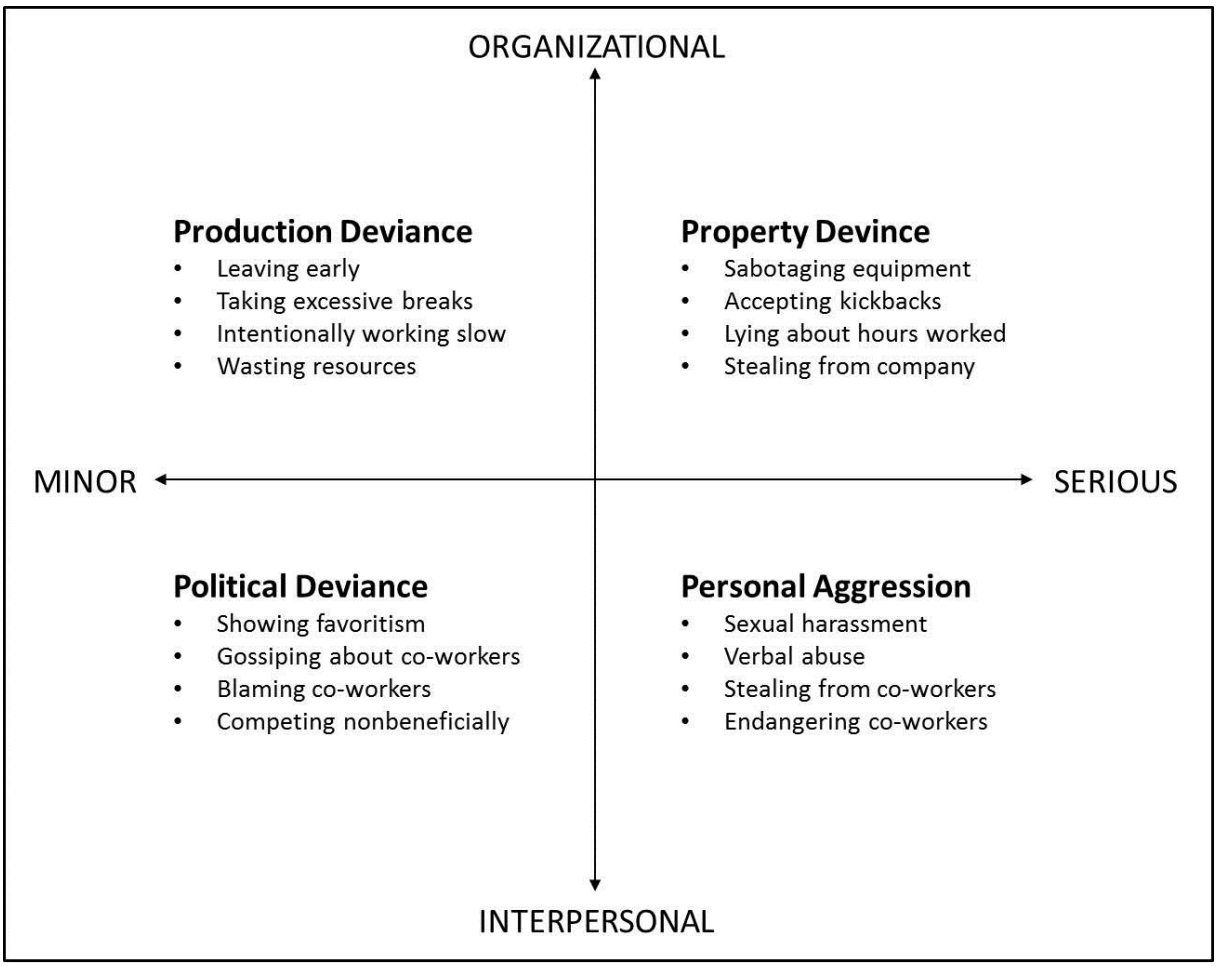

Figure 4. Classification of CWB by Robinson and Bennett. From "A typology of deviant workplace behaviors: A multidimensional scaling study," by S. L. Robinson and R. J. Bennett, 1995, *Academy of Management Journal, 38*, p. 565.

Third, Gruys and Sackett (2003) proposed an eleven-facet structure of CWB and investigated the dimensionality of CWB guided by the questions of whether different forms of CWB frequently occur together. The authors collected relevant behavior from the literature and formed it into categories according similarity. Overall, they found 66 behaviors (e.g., lying to an employer or supervisor to cover up a mistake) and 11 categories of CWB (e.g., misuse of information). Participants were asked to rate the likelihood of engaging in each act of CWB; afterward these data were transformed into a matrix of similarity that conducted multidimensional scaling analysis.

The results revealed that CWB typically co-occurs. Moreover, Gruys and Sackett found a two-dimensional solution for the classification of CWB. The first dimension refers to the target of CWB (interpersonal vs. organizational), which is in line with Robinson and Bennett's (1995) results. The second dimension refers to the relevance of CWB concerning the task of the perpetrator's job (relevant vs. non-relevant task). CWB with low relevance for performing a job task could include inappropriate verbal or physical actions, whereas CWB with high relevance for performing a job task could be the misuse of time and resources.

The fourth and final widely known structure of CWB was developed by Spector et al. (2006). Based on items from the 45-item Counterproductive Work Behavior Checklist (Spector et al., 2004), their five-dimensional approach includes the CWB categories of abuse, production deviance, theft, sabotage, and withdrawal. The authors' classification was guided by the idea that different types of CWB tend to have different potential antecedents. For example, abuse tends to be associated more with job stressors and upsetting emotions than with psychological strains or boredom. Moreover, the five categories were also classified into interpersonal CWB, which consists mainly of abuse items, and organizational CWB, which consists mainly of items from the other four categories.

In conclusion, there is still no full agreement among researchers on the internal structure of CWB—that is, how the broad array of CWBs are related to one another and how they should be categorized (Marcus, Taylor, et al., 2016). Further research indicated that a hierarchical model of CWB might exist, as exemplified in Sackett and DeVore's (2001) three-level model. According to this approach, the highest level of CWB comprises a general factor of overall CWB, whereas the level below contains group factors of CWB such as deviance towards individuals or organizations. Finally, the lowest level consists of individual

behaviors (such as theft). Moreover, the authors proposed a common latent factor that underlies all forms of CWB (Sackett & DeVore, 2001).

Determinants of CWB

In addition to the questions regarding the structure of CWB, approaches concerning the reasons of CWB emerged (e.g., Spector & Fox, 2002). There are three main empirically justified variables that influence the occurrence of CWB (Nerdinger et al., 2008): (a) (perceived) injustice (e.g., Ambrose, Seabright & Schminke, 2002; Greenberg, 1990) in conjunction with frustration (e.g., Fox, Spector, Goh & Bruursema, 2007; Spector, 1997), personality traits such as (b) conscientiousness (e.g., Berry, Ones & Sackett, 2007; Dalal, 2005), and (c) self-control (e.g., Marcus & Schuler, 2004). Unlike the first variable, the latter two dispositional variables have been discussed before in conjunction with integrity. Moreover, regarding personality traits, the conglomerate of conscientiousness, emotional stability, and agreeableness that has been found in relationship with integrity is also in relationship to CWB (Berry, Ones & Sackett, 2007). In line with integrity, conscientiousness has the strongest negative relationship with CWB (Sackett & DeVore, 2001).

In a meta-analysis by Dalal (2005), further important correlates of CWB were introduced: Dalal (2005) found a corrected mean correlation of $\rho = -.37$ between job satisfaction and CWB. Furthermore, commitment had a corrected mean correlation of $\rho = -.36$ to CWB.

Another category of organizational behavior that was included in Dalal's (2005) meta-analysis is organizational citizenship behavior (OCB). OCB refers to voluntarily completing extra work for the organization (e.g., improving work processes) or coworkers (e.g., helping coworkers) beyond the formal requirements of the job position (Organ, 1988, 1997). A negative moderate corrected mean correlation of $\rho = -.32$ was found between CWB and OCB (Dalal, 2005). In addition, CWB as well as OCB are both predicted by similar non-personality factors and by similar personality factors (e.g., Bowling, 2010). Moreover, CWB reveals a theoretical connection to OCB (Borman & Motowidlo, 1993). Nevertheless, CWB and OCB are not opposite poles of one dimension.

In summary

CWB is a clearly defined construct. Regarding its structure, a vast number of CWBs can be structured into different categories, and CWB is therefore an overall construct with specific deviant behavioral domains. All CWBs are positively intercorrelated (Marcus & Schuler, 2004; Sackett & DeVore, 2001). The correlative determinants in the nomological network of CWB are congruent with the determinants of integrity. This fact underlines the narrow oppositional relationship between integrity and CWB.

2.4 FAKING ON INTEGRITY TESTS

The desire to make a good impression is an intrinsic element of the human experience. If we meet a person of interest (e.g., a future in-law) or want to achieve something (e.g., to get a job), we aim to make a good impression. This phenomenon is referred to as social desirability, and its theoretical construct can be divided into two main categories: deceiving oneself in terms of one's own attributes (self-deception) and enhancing one's own characteristics in order to make a good impression on others (impression management) (Paulhus, 1984). While self-deception is an unconscious process, impression management is used consciously and for strategic purposes.

The concept of faking is similar to impression management but also entails certain differences: Impression management involves the attempt to impress others in daily situations, whereas Berry and Sackett (2009) defines applicant faking as a process that occurs in addition to everyday impression management in the context of personnel selection and with the goal of being hired for a job. In applicant faking, applicants tend to put a positive spin on their answers or behavior in order to increase the appearance of preferable personality characteristics and diminish the appearance of negative characteristics (e.g., Gerber-Braun, 2010; Griffith et al., 2007).

Studies have shown that different methods of personnel selection are vulnerable to this distortion by participants (e.g., McFarland, Ryan, & Kriska, 2003; Tett, Freund, Christiansen, Fox, & Coaster, 2012). Integrity tests target socially desirable behavior, attitudes, and traits and are therefore likely to be faked by test-takers (Dwight & Alliger, 1997; Karren & Zacharias, 2007). Indeed, these tests are among the most faked methods of

assessment (McFarland, & Ryan, 2000). Moreover, integrity tests are faked by participants in both anonymous (Marcus, 2006) and simulated selection settings (Gerber-Braun, 2010; Jackson, Wroblewski, & Ashton, 2000) as well as by applicants in personnel selection procedures (Birkeland et al., 2006).

Models of faking

Various theoretical models of faking behavior and of the psychological processes that underlie faking have been developed. The three most well-known models are presented in the following section (Goffin & Boyd, 2009; McFarland & Ryan, 2000, 2006; Snell, Sydell, & Lueke, 1999).

The model of faking proposed by Snell et al. (1999) is based on two main components: the ability to fake and the motivation to fake, both of which affect a person's tendency to fake. The first component—the ability to fake—refers to the ability of test takers to distort their responses and is influenced by variables such as dispositional factors, experiential factors, and test characteristics. The second component—the motivation to fake—refers to a person's intention to fake and is influenced by demographic factors, dispositional factors, perceptual factors, and contextual factors.

McFarland and Ryan's (2000) faking model concentrates on variance in faking with regard to non-cognitive measures. Different individual variables (e.g., values, morals, personality traits) influence beliefs about faking, and these beliefs affect the intention to fake. This relationship between beliefs about faking and the intention to fake is influenced by situational variables (e.g., the desire for a job), which act as moderators in the relationship between beliefs and the intention to fake, with the intention to fake resulting in the behavior of faking. This relationship between the intention to fake and faking behavior is influenced by two moderators: the ability to fake (e.g., item transparency) and the opportunity to fake. In 2006, McFarland and Ryan (2006) expanded their model by integrating the theory of planned behavior, which includes three components: attitudes, subjective norms, and perceived behavioral control. The authors suggest that these components should predict faking by influencing a person's intention to fake.

A third, widely known general model of faking was developed by Goffin and Boyd (2009). Much like Snell et al.'s (1999) model, Goffin and Boyd's (2009) model focuses on two components: the ability to fake and the motivation to fake. Both components of the model

are influenced by different individual and contextual antecedents (e.g., personality traits, perceived opportunity to fake). Nevertheless, in contrast to the two previous models of faking, Goffin and Boyd's (2009) model includes changes regarding some main points: (a) The focus of the model lies on the item level instead of on a complete personality test. (b) The component of ability is defined more subjectively as the perceived ability to fake. (c) The authors emphasize the importance of clarifying the relationship between the perceived ability and motivation to fake and suggest that the perceived ability to fake influences motivation. The perceived ability to fake thus has a direct causal effect on the motivation to fake. (d) Goffin and Boyd (2009) propose a more exhaustive list of factors that refer to individual differences or contextual antecedents and directly influence a person's motivation or perceived ability to fake.

Finally, all three models of faking are based on a person's ability to fake as well as on his or her motivation or intention to fake, but they also differ with regard to the relationship between both components. Snell et al.'s (1999) faking model defines the ability and motivation to fake as two separate constructs that directly influence faking. In contrast, McFarland and Ryan (2000) find that the ability to fake influences the relationship between the motivation (intention) to fake and faking behavior. Moreover, Goffin and Boyd (2009) take a direct causal effect of perceived ability on the motivation to fake as a basis in contrast to two separate constructs. Moreover, in contrast to the two existing models, Goffin and Boyd (2009) assume that there is a direct causal effect of the perceived ability to fake on the motivation to fake. Moreover, Goffin and Boyd (2009) define the component of ability more subjectively as a *perceived* ability to fake.

McFarland and Ryan (2000) include the concept of beliefs about faking in their faking model and assume both that these beliefs directly influence the motivation to fake and that this relationship is influenced by several variables. In line with these influencing variables, Goffin and Boyd (2009) propose two exhaustive lists of factors relating to individual differences or contextual antecedents. The variables of one of these lists directly influence a person's motivation, and the variables of the other list directly influence the perceived ability to fake. In general, the focus of Goffin and Boyd's (2009) model lies on the item level instead of on an entire personality scale.

Consequences of faking

Faking leads to three possible effects on diagnostic assessment methods in general and on integrity tests in particular:

(a) The scores of a measure can change. This means, for example, that participants with a low or average tendency of a characteristic receive a high score for this characteristic on the measure. McFarland and Ryan (2006) demonstrated that an integrity test score is significantly higher in the faking condition than in the honest condition. One meta-analysis found that personality traits are influenced by applicant faking, with an effect size ranging from $d = .11$ to $d = .45$ (Birkeland et al., 2006). Moreover, another meta-analysis revealed that faking increases scores on integrity tests by about one-half and up to one standard deviation depending on the kind of integrity test used ($d = .59$ for personality-oriented integrity tests, $d = 1.02$ for overt integrity tests) (Alliger & Dwight, 2000).

(b) The validity of the measures can be threatened. This means that faking may have an impact on test results by rendering them invalid. In a review by Ones and Viswesvaran (1998b), social desirability—the theoretical construct underlying faking—was not found to affect the criterion-related validity of integrity tests. Further studies have confirmed this finding with regard to personality tests in general (e.g., Ones, Viswesvaran & Reiss, 1996). Moreover, studies found little impact on the construct validity of personality measures (e.g., D. B. Smith & Ellingson, 2002). In contrast to this finding, still other studies have revealed that faking does indeed affect the test validity of selection methods (e.g., Griffith & Peterson, 2008; Jeong, Christiansen, Robie, Kung, & Kinney, 2017; Mueller-Hanson, Heggestad & Thornton, 2003; Tett & Christiansen, 2007).

(c) The rank order of qualified applicants can change. This means that participants with lower levels of a desirable personality trait sometimes claim to have a higher level and thus achieve better test results than other participants whose actual level of this desirable personality trait is higher (e.g., Anglim, Lievens, Everton, Grant, & Marty, 2018). Moreover, the amount of faking is not constant and varies as a function of situational factors (e.g., context, subscales; Birkeland et al., 2006) and personnel factors (e.g., respondents' perceptions of social pressure, their beliefs concerning the ease of faking; McFarland & Ryan, 2006). Depending on these factors, some applicants fake their responses to a greater extent than others, thereby undermining participants' correct rank order (e.g., Anglim et al., 2018; Birkeland et al., 2006; Peterson et al., 2009). As a result, applicants who are not the

most suitable employees for a job may be hired (Griffith et al., 2007; Rosse, Stecher, Miller, & Levin, 1998).

Methods to decrease faking

Different approaches have been developed to decrease faking in the field of integrity tests and other diagnostic assessment methods related to more sensitive topics (e.g., income, sexuality) (see Table 4). A common feature of integrity tests is the use of validation subscales or special scales to detect applicants' answer distortion (Crowne & Marlowe, 1960; D. G. Fischer & Fick, 1993). For both methods, researchers agree that scales cannot effectively indicate the variance caused by participant faking (e.g., Griffith & Peterson, 2011; Peterson, Griffith, Isaacson, O'Connell, & Mangos, 2011; Sackett, 2011).

To control for faking, measures that make use of time latency could be used in addition to traditional lie scales (Dwight & Alliger, 1997; Fine, & Pirak, 2016). For example, the IAT for workplace integrity is an implicit measure of integrity that consists of three related IAT categories that assess personal (self/other), social (employee/employer), and attribute (honesty/dishonesty) contrasts (D. Fischer & Bates, 2008). Analyses are based on differences between reaction times: People who fake take more time to answer than do people who do not fake.

With regard to the test design, a simple method to combat faking could be to provide a warning in the test instructions indicating that faked responses can be identified by checking the answer structure (McFarland, 2003). In addition, test formats (e.g., situational judgment tests) or rating formats (e.g., forced choice) can reduce socially desirable answering (Jackson et al., 2000; McDaniel, Hartmann, Whetzel, & Grubb, 2007). With regard to test analyses, new methods have also been developed, such as the ipsatization procedure (i.e., item responses are standardized for each participant) and partial correlation (i.e., the response set is psychometrically controlled).

New physiological methods, for example eye tracking, may be used to prevent faking in the near future. One study revealed that participants in the *fake good* condition were slower to answer and displayed more eye fixations on both extreme answering options (Van Hooft & Born, 2012). Since faking is a cognitive process that requires mental resources, slower reaction times or more eye fixations can indicate faking. The use of physiological as well as neurological methods—such as fMRI—to uncover faking on integrity tests is thus

promising; however, due to legal and ethical concerns, these methods may not yet be applied. Moreover, additional studies are needed to investigate the mechanism underlying these methods as well as their validity.

While many methods aim to reduce faking, according to Burns and Christiansen (2011), no effective method yet exists that prevents faking to a satisfactory level. Moreover, researchers have described applicant faking as "apparently the most pervasive concern that personnel selection practitioners and researchers alike have regarding personality assessment […]" (p. 152, Goffin & Boyd, 2009).

Faking also remains a problem on integrity tests and can cause an obvious shift in test values (Allinger & Dwight, 2000). Moreover, overt integrity tests are more susceptible to faking (e.g., Alliger & Dwight, 2000). Researchers thus recommend searching for new methods to effectively prevent faking on integrity tests (Karren & Zacharias, 2009).

Table 4

Overview of methods to combat response faking

Method	Explanation of the method	Prevents vs. detects faking	Advantage(s) of the method	Disadvantage(s) of the method
Anonymity	Respondents are offered anonymity.	prevents	simple and fast widespread	can stimulate faking
Bogus pipeline technique	Respondents are told that untruthful answers can be detected.	prevents	simple and fast, intended to improve the truthfulness of self-reports	unethical for participants, doubts exist about its effectiveness
Confidentiality	Researcher provides explicity assurances to participants that their answers will remain completely confidential.	prevents	simple and fast	opposite effect due to the lack of verification of the truthfulness, can stimulate faking
Counterbiasing method	Counterbiasing information is provided to the respondents and makes them believe that the less desirable behavior may be the norm.	prevents	simple and fast	doubts exist about its effectiveness, different results evoke due to the type of reference and the wording of frequency
False consensus effect	People tend to overestimate the degree to which their own attitudes or behaviors represent the norm for others.	prevents	theory-based element, highly statistically significant and a moderate effect size[a]	disagreement exists about the underlying mechanism of the effect

Method	Explanation of the method	Prevents vs. detects faking	Advantage(s) of the method	Disadvantage(s) of the method
Forced-choice technique	A single option must be selected from multiple given answers.	prevents	empirical evidence is given	provides relative—albeit not absolute—answers
Implicit goal priming	The goal of honesty is primed in respondents before they fill out a questionnaire.	prevents	empirical evidence is given, well-tested in laboratory settings	priming may not be strong enough or fail, concerns exist about priming outside the laboratory
Indirect questioning	Respondents answer items from the perspective of some ambiguous, average other person.	prevents	promising results for sensitive constructs	validity of the inferential leap: the only assumption is that participants respond about themselves
Ipsatization	Item responses from each participant are standardized and compared with those of other participants by relating each answer to the individual's own mean.	detects	the ipsative format has the power to reduce bias, thereby enhancing validity	difficulties exist in making comparisons between people; statistical values calculated from ipsative test scales (e.g., means, correlations) have to be interpreted with care because they are not independent; caution is needed when using ipsatized data in factor-analytic designs
Monitoring	The answer bias is visually controlled.	detects	simple and fast	is not systematic, doubts exist about its effectiveness
Partial correlation	The response set is psychometrically controlled.	detects	simple and fast	diminishes criterion-related validity if relevant variance is removed

Method	Explanation of the method	Prevents vs. detects faking	Advantage(s) of the method	Disadvantage(s) of the method
Randomized response technique	The sensitive question is worded as two dichotomous alternatives; one question must be answered honestly, but the researcher does not know which; the honest answer thereby remains unknown, but can be determined.	detects	anonymity can be created credibly, empirical evidence is given[b], includes a range of methods (e.g., unmatched count technique)	does not yield individual-level data, there is a tendency of this method to fail in online mode
Response-time measure	The differences in reaction time are analyzed.	detects	simple and fast	doubts exist about its effectiveness
Situational judgement tests	Situational judgement tests normally present a range of different situations with a multiple-choice response format; these tests are very popular for recruitment.	prevents	can cover the construct, which is very well investigated	complex test construction, doubts exist about their validity
Social desirability scale	A scale is used that measures socially desirable responding.	detects	empirical evidence is given	scale scores do not function as a moderator, suppressor, or mediating variable; participants must spend extra time answering the questionnaire, spuriousness effect by correlation with dependent or independent variable

Method	Explanation of the method	Prevents vs. detects faking	Advantage(s) of the method	Disadvantage(s) of the method
Unmatched count technique[c]	Two groups of participants respond to a handful of non-sensitive items; for one group there is an additional sensitive item. The mean number of agreements from both groups is compared to estimate average responses to the sensitive item.	detects	anonymity can be created credibly	technique only estimates the proportion of an undesirable behavior on a group level instead of for each individual; relative loss of data points inflates sampling error
Warnings (pre- or middle-warning)	A warning is given at the beginning or in the middle of a test indicating that untrue answers can be detected.	detects	simple and fast	might lead to distrust in the measure among honest applicants

Note. [a] see meta-analysis by Mullen et al. (1985); [b] see meta-analysis by Lensvelt-Mulders, Hox, van der Heijden, & Maas (2005); [c] also called the item count technique, the unmatched block design, or block total response

In summary

Faking is a conscious process of response distortion by applicants. Models of faking include the ability and motivation to fake and are influenced by different individual and contextual factors. The consequences of faking are vast in terms of non-cognitive, sensitive measures—especially on integrity tests—and can lead, for example, to an altered ranking of applicants, thereby threatening the criterion and construct validity of integrity tests altogether. To reduce faking, many methods that control for or prevent faking have come into use, such as social desirability scales, statistical approaches, methods based on psychological effects or biases (e.g., the false consensus effect), and new test formats (e.g., situational judgement tests). Nevertheless, none of these methods has proven capable of effectively diminishing faking.

CHAPTER 3 REVIEWING INTEGRITY TESTS

Integrity is a well-known concept and is measured via integrity tests. Many such tests exist with a vast array of different scales and construction approaches. To obtain an overview and detailed insights into existing measures, this chapter begins with an overview of 76 existing integrity tests and compares the test scales of 50 such tests. Within this chapter, 16 tests are examined in detail and compared in terms of different aspects of test construction (e.g., item format, internal consistency) and test application (e.g., test aim, validity). Finally, the key findings of all analyses are discussed in greater detail, and recommendations for test uses are included. Based on all analyses in this chapter, an appropriate test for subsequent empirical data collection in this dissertation is chosen.

3.1 OVERVIEW AND SCALE ANALYSES OF INTEGRITY TESTS

The first publication to present an overview of integrity tests was published in 1989 (O'Bannon et al., 1989). In their handbook, the authors presented 43 integrity tests with detailed test information. Shortly after this publication, Snyman et al. (1991) published an overview of 16 integrity tests in which the authors reported findings on the reliability, validity, and fakability of eleven overt (e.g., Compu-Scan) and five personality-based (e.g., Personnel Reaction Blank) tests. Coyne and Bartram (2002) presented a structured framework of six aspects to assess the quality of integrity tests consisting of scope, accuracy (referring to reliability), relevance (referring to validity), fairness, acceptability, and practicality. However, Coyne and Bartram (2002) only mentioned integrity tests exemplarily and did not systematically compare the same integrity tests in terms of all six aspects.

Regarding these approaches to listing integrity tests, Ones and Viswesvaran (1998b) reported 45 commercial integrity tests that were available in the United States at the end of the last century. Since then, the test market has continued to grow, thereby rendering an overview increasingly difficult and begging the question of the current number of integrity tests and their focus regarding the construct of integrity.

To identify existing integrity tests, all articles and book chapters in the databases PsycINFO and PSYNDEX that deal with at least one integrity test were gathered[4]. The

4 Data were collected in the period from 2011 to 2014.

keywords used were *integrity test* and *honesty test*. Moreover, the 9th to the 18th Volumes of the Mental Measurements Yearbook (MMY) series published by the Buros Institute of Mental Measurements was screened. The MMY is a book series that contains comprehensive collection of commercially available measures from 18 areas of psychological testing. For the MMY series, the keywords used were *integrity*, *(dis-)honesty*, *reliability*, *trustworthiness*, *deviance*, and *theft*. In addition to the literature search, test publisher websites were scanned for further information. Test manuals were available online for some integrity measures.

The search for integrity tests was met with several challenges that had previously been identified (e.g., Camara & Schneider, 1994; Van Iddekinge et al., 2012a): (1) Integrity test items are sometimes removed from one measure and formed into a new measure. (2) Integrity tests are sometimes renamed due to revisions. (3) Integrity test naming does not always appear to refer to integrity or integrity-related expressions. (4) Test information is occasionally unavailable as test publishers do not always offer information about their measures. (5) In some cases, test information is reported incorrectly. This overview thus does not claim to be exhaustive regarding these five issues.

In total, 76 integrity measures were identified (see Appendix, Table A). A test was taken into account as an integrity test if it (a) predicts narrow or broad CWB in a person and (b) was developed for the purpose of screening applicants or employees (Association of Test Publishers [ATP], 2010). Moreover, borderline cases of personality tests that are used in personnel selection processes and that comprise one or more subscales with reference to integrity were also included (e.g., the Hogan Personality Inventory from R. Hogan, 1986). On the other hand, conditional reasoning tests that measure specified kinds of CWB—such as aggression (e.g., the Conditional Reasoning Test - Aggression from James & McIntyre, 2000)—were not included.

The majority of the identified integrity measures stem from North America, with only a few measures originating in Europe or elsewhere (e.g., the IntegriTest from Midot, 2007, comes from Israel). Many measures are provided in more than one version (the maximum was the Personnel Selection Inventory from Press, 1980, with 20 versions). Many measures were developed prior to 2000 (86%), with only ten tests (15%) having been developed since 2000. There were 54 overt tests (75%), 10 personality-based tests (14%), and 8 hybrid tests (11%). These findings are in line with the historical development of integrity tests, which

were rapidly developed in North America during the end of the 1980s and the beginning of the 1990s. A clear predominance of overt integrity tests could be found, which indicates that the construct of integrity is more strongly defined by a conglomerate of attitudes, behavioral tendencies, and convictions instead of personality traits.

A qualitative and quantitative analysis of subscales of integrity tests was conducted (*N* =50; 26 additionally listed integrity tests were not taken into consideration because information about the number and naming of their subscales was not available in open sources). The qualitative analysis of integrity test subscales revealed a heterogeneous range of characteristics: Subscales refer to individual elements, such as personality (e.g., conscientiousness, emotional stability), attitudes (e.g., work ethics, drug avoidance), abilities (e.g., cognitive ability), or other individual facets (e.g., burnout, service orientation). Further subscales involve conditions of employment (e.g., tenure, part-time contract) or check for participants' honesty when answering (e.g., candidness, accuracy). This variety of subscales emphasized the notion that integrity tests should be carefully classified into one homogenous measure as they currently appear to be a very heterogeneous measure. Moreover, as some subscales belong to other constructs (e.g., burnout, accuracy), not all subscales may be relevant for the construct of integrity or reflect its basic nature.

With regard to a quantitative analysis of integrity test subscales, an average of six subscales (*SD* = 4) per test were identified, with a minimum of 1 and a maximum of 26 subscales. The vast range of subscales supported the missing key concept of integrity. In addition, large numbers of subscales indicated that features of other constructs might be involved. The most common subscales were substance abuse/resistance (54%), (dis-)honesty (42%), and validation (40%). A validation subscale meant that a test included a protection mechanism against providing socially desirable answers. In most cases, a person was deemed to provide socially desirable answers if he or she agreed very often with socially desirable items (e.g., "I'm always on time."). Focusing on the criteria that integrity tests aim to predict, the most common CWB subscale was theft (19%). Other common CWB-related subscales were safety orientation (15%) or absenteeism (12%). The most widespread subscales indicated that there was a common basis of the nature of integrity. Nevertheless, these subscales were part of other constructs (e.g., validation scales referred to the construct of social desirability and substance abuse was part of the CWB concept).

In summary

This section aims to provide an overview of existing integrity tests and to analyze the subscales of these tests. In line with the predominance of overt tests, the variety of subscales indicates a missing clear boundary for the construct of integrity. Moreover, regarding the content of subscales, which often refers to other constructs, integrity seems to act as a general container for every feature that is somehow related to it. As a consequence, integrity tests should be carefully classified into a homogenous category. With regard to scientific needs, results of a study on integrity should not be interpreted without considering the used integrity measure. Moreover, focusing on integrity tests more individually is recommended to efficiently evaluate the tests, their results, and their research findings. The test quality may thus be improved, and the general validity may be enhanced for test application in both research and practice.

3.2 ANALYZING INTEGRITY TEST CONSTRUCTION

A selection of tests based on the overview of integrity tests presented in the previous section was taken for detailed systematic analyses in Chapters 3.2 and 3.3. The criteria used in choosing a test were novelty, publicity, and variety. The first criterion, “novelty”, referred to test publication since the year 2000, and seven tests met this criterion. The second criterion, “publicity”, included frequent citations and use of a test in studies, in the review by O'Bannon et al. (1989), in the MMY series, or in a list of tests within the appendix of a book chapter or articles (e.g., Cullen & Sackett, 2004; Ones & Viswesvaran, 1998b). Again, seven tests met this criterion. The third criterion, “variety”, comprised different characteristics of integrity tests regarding elements such as different types of integrity tests (i.e., overt, personality-based or mixed-type), different response formats (e.g., agree/disagree, multiple-, or forced choice), or different target populations (i.e., applicants, employees, or both groups). Two further measures of integrity were thereby included. Finally, only tests for which sufficient information was obtained were included, and some well-known integrity tests could therefore not be considered. In summary, 16 integrity tests were selected. These tests are presented in detail and systematically compared (see Table 5).

The construction of a test begins with the choice of a strategy; therefore, the first criterion is the strategy of development, which is primarily criterion-orientated without any underlying theory or model. Nearly all integrity tests are thus developed via a literature search either to generate an item pool that is reduced by statistical testing (internal test construction strategy) or to focus on empirically relevant correlates of integrity (external test construction strategy). In some cases, both strategies are mixed. For example, the manual of the Applicant Risk Profiler describes a literature review that refers to personality traits that correlate with CWB (e.g., theft or drug use) and with former behavioral indicators (e.g., aggressive tendencies and attendance issues) (Llobet, 2009). In some cases, tests constructions have been guided by specific parts of theories. For example, the Stanton Survey manual describes that the test is guided by three assumptions that relate to the differential association theory of Sutherland (1940) and to the concept of rationalization (Sykes & Matza, 1957; Matza, 1964). One of the sixteen tests is based on a theoretical framework. The Giotto is based on a model of personality originally developed by Latin literary figure Aurelius Prudentius Clementis (Rust, 1999). In addition to the finding of a lacking theory and regarding to the definition of the integrity construct, which should also underlie every integrity test, no definition could be found in any avaliable test manual.

The next criterion of test construction is the size of the norm sample, which ranged from 191 participants for the Counterproductive Behavior Index to 45,000 participants for the Hogan Personality Inventory–Reliability Scale. With regard to the primary application of integrity tests in personnel selection, a norm sample of at least 300 participants is essential (Evers, 2001). All norms except for two measures (CBI, IBES) were adequate in size. Upon more closely examining the categories of the norm sample, only two available tests manuals had norm categories referring to gender, age, education, or characteristics of work. The Counterproductive Behavior Index had norm categories regarding age and education, and the Inventory of Job-Related Attitudes and Self-Assessments included norm categories regarding gender, age, work experience, and the industrial sector. Instead of a norm-reference, some integrity tests referred to a criterion-reference with regard to certain cut-off scores. For some measures, no norms were available in empirical studies or free test manuals.

Table 5

Aspects of test construction for 16 integrity tests

Test acronym	Strategy of test development	Size of norm sample	Type of test	Number of subscales	Validity subscale	Number of items	Response format	Internal consistency
API	internal	no	overt	10	yes	64	agree/disagree	.74
ARP	internal & external	3,362	overt	5	yes	65	agree/disagree	.79
CBI	internal	191	overt	8	yes	140	agree/disagree	.94
ERI	external	43,000	both	7	yes	81	true/false	n/a
ESQ	external	n/a	personality-based	15	no	27	forced choice	.83
Giotto	model	701	personality-based	7	no	101	vary	.73
HPI-R	contrast	45,000	personality-based	4	no	18	true/false	.75
IntegriTEST	n/a	5,000	overt	4	no	96	yes/no/not known	.84
IBES	external	232	both	9	no	115	agree/disagree	.93
IPI-R	rational	2,385	personality	26	yes	310	true/false	.52
PRB	external	558	personality-based	4	yes	84	vary	.65/.97
PSI	n/a	n/a	overt	11	yes	var	vary	n/a
Reid Report	n/a	n/a	overt	6	no	61	vary	.90
Stanton Survey	n/a	466	overt	3	no	83	vary	.91/.93
ViewPoint (W6)	external	1,100	overt	8	yes	81	multiple choice	.90/.95
Workkeys PA	external	692	both	2	no	60	agree/disagree	.79/.90

Note. API = Applicant Potential Inventory, ARP = Applicant Risk Profiler, CBI = Counterproductive Behavior Index, ERI = Employee Reliability Inventory, ESQ = Employee Screening Questionnaire, HPI-R = Hogan Personality Inventory–Reliability Scale, ITEST = IntegriTEST, IBES = Inventar berufsbezogener und einstellungsbasierter Selbstbeurteilung [Inventory of Job-Related Attitudes and Self-Assessments], IPI-R = Inwald Personality Inventory–Revised, PRB = Personnel Reaction Blank, PSI = Personnel Selection Inventory, Workkeys PA = Workkeys Performance Assessment.

"n/a" = information was not available, both = a test refers to an overt as well as a personality-based typ.

Internal = literature search. External, rational or contrast = correlates of integrity. Var = Number of items is varying.

Regarding the criterion of the type of tests, eight measures referred to overt integrity tests, four measures referred to personality-based integrity tests, and three measures assessed both kinds of integrity tests. One measure (the Inwald Personality Inventory–Revised) was described as a personality inventory with a special focus on selecting police and security personnel. By contrast, the Reliability Scale of the Hogan Personality Inventory is a part of the personality inventory and described as a personality-based measure.

In order to compare test construction principles in more detail, the subscales of tests were further characterized by two criteria: the number of subscales and the existence of a validity subscale. The number of subscales ranged from 2 (for the Workkeys Performance Assessment) to 26 scales (for the Inwald Personality Inventory–Revised) with a mean of eight scales (*SD* = 6). The number of discovered test criteria was comparable to the scale analysis in Chapter 3.1. As mentioned in the section before, this large and varying number of subscales highlighted not only the multidimensionality but also the lack of clarity of the integrity construct. According to validity subscales, about half of the selected integrity tests included a validity subscale to detect socially desirable responses (e.g., the ViewPoint, version W6). Moreover, the measures additionally included subscales to identify further distortions such as random answering or response sets (e.g., the Applicant Potential Inventory).

More specifically, tests were further explored via the number of items, their respone format, and their internal consistency. Regarding the number of items, tests ranged from a minimum of 27 items for the Employee Screening Questionnaire to a maximum of 310 items for the Inwald Personal Inventory–Revised, with a mean of 93 items (*SD* = 63). In line with the vast number of subscales, the number of items was large. In addition to this large and strongly varying number of items, the response format of items often differed. That is, the response format appeared as agree/disagree, true/false, yes/no, multiple choice, or forced-choice. For example, the Applicant Potential Inventory has a 4-point answer format for respondents to agree or disagree with a statement, while the ViewPoint (W6) has a multiple-choice answer format. Moreover, one third of integrity tests additionally change their response format within the test or within different test versions (e.g., the Personnel Reaction Blank). According to the test construction literature (e.g., Lienert & Raatz, 1998), changing the response format within a personality-related test should be prevented.

The final criterion of internal consistency of integrity tests ranged from .65 to .97 between tests. In line with this finding, similar values were found for retest reliability: Scores ranged from .58 to .97 among tests. As integrity tests are primarily applied in personnel selection settings, tests with a reliability of at least .80 are classified as acceptable, and tests with a reliability greater than .90 are classified as good (Evers, 2001). In the case of internal consistency, four tests were acceptable and five tests were good. Five tests were not acceptable with regard to their internal consistency, and data on internal consistency were not available for two tests.

In summary

Sixteen integrity tests were identified for this comparison. In general, these kinds of tests are far from homogeneous (Murphy, 2000). With regard to their construction, a pervasive lack of theory is evident because most integrity tests are a-theoretically developed, and there is no definition of the underlying construct of integrity. In addition, integrity tests come in many varieties (e.g., content of subscales, number of items). As a consequence, this kind of test has no clear focus. Moreover, ensuring the quality of the tests (e.g., size of norm sample, reliability scores) is paramount, and the tests should therefore maintain a clear focus with a smaller number and a more precise content of subscales.

3.3 ANALYZING INTEGRITY TEST APPLICATION

Based on the selected 16 integrity tests of Chapter 3.2, a detailed examination of integrity tests' application is necessary to complete their analyses (see Table 6). The first step is to focus on the definition of tests that are used to detect future employees who tend to display CWB (ATP, 2010; Ones & Viswesvaran, 1998b). In line with this definition of integrity tests, their primary target population is applicants. As expected, the majority of 14 tests were designed for applicants. One measure (the Hogan Personality Inventory–Reliability Scale) is intended for employees, and the Personnel Reaction Blank is intended for both, applicants and employees.

Another criterion of the definition of integrity tests is that they forecast the tendency of displaying CWB. It is therefore important to consider the aim of the 16 selected integrity

tests in detail. In contrast to the classical definition of integrity tests (which is to identify deviant job applicants), these tests additionally aimed at detecting the performance of job applicants. This criterion is in line with research results, which demonstrate that integrity tests predict job performance (e.g., Ones et al., 1993; Van Iddekinge et al., 2012a). However, this is not in line with the definition of integrity tests. Subscales thus often refer directly to CWB and therefore muddle the construct of CWB.

To compensate for the differences in integrity test subscales, most test manuals recommend interpreting the overall score instead of single subscale scores. Integrity tests, which are usually multi-dimensional and comprise many CWB-related subscales, are not based on any theory or construct. Therefore, the overall score represents a stable tendency of the test taker to display behavioral integrity.

As to an efficient application, it is important to assess how long it takes test users to complete the measure. The duration of a test (reported in minutes) was therefore included as a criterion. The criterion of duration in minutes is more precise than that of the number of items because it takes the response format into account. For example, the Employee Screening Questionnaire includes 27 items with a more complex answer format of forced choice and takes 20 minutes. In contrast, the IntegriTest follows the same path but includes 96 items with a simple “yes/no/not known” response format. With regard to the range of duration, the data reveal that a minimum of 5 minutes and a maximum of 45 minutes are necessary to complete the Hogan Personality Inventory–Reliability Scale and the Inwald Personal Inventory–Revised, respectively. The average time to complete an integrity test is about 20 minutes (SD = 10). It is necessary to consider that some time specifications vary across publications or with reference to the test version (e.g., for the Personnel Selection

Table 6

Criteria of test application for 16 integrity tests

Tests	Purpose population	Aim of test	Duration [minutes]	Web-based format	Language	Free manual	Validity (CWB)[2]
API	applicants	to select high-performing employees to reduce their CWB	15	yes	E, F, S	no	-.60
ARP	applicants	to identify a potential risk of applicants for their interactions with other members of the organization or for themselves	< 30	yes	E	yes	-.47
CBI	applicants	to identify job applicants who are deviant or have an extreme behavior or attitude in one or more of six work dimensions that influence productivity	15	yes	E	yes	n/a
ERI	applicants	overall measure of the likelihood that an applicant will perform reliably and productively on the job	12-15	no	E, F, S	yes	n/a
ESQ	applicants	to predict broad criteria of CWB as well as to predict positive job behavior; to detect good-, high-quality, dishonest, and unproductive employees	20	yes	n/a	no	-.47
Giotto	applicants	to assess a wide range of integrity-related personality	5-20	no	E	yes	-.38
HPI-R	employees	to detect honest and reliable employees who are “responsive to supervision”	5	yes	various	yes	-.45
IntegriTEST	applicants	designed to predict deviant work behavior, such as theft, fraud, drug use and bribery among job applicants and employees in selection scenarios	20	yes	various	no	-.25

Tests	Purpose Population	Aim of test	Duration [minutes]	Web-based format	Language	Free Manual	Validity (CWB)[1]
IBES	applicants	to predict a specific behavioral criterion that represents CWB at best	15-25	yes	G	no	-.49
IPI-R	applicants	to identify appropriate applicants in public safety, security, or law enforcement for an application after a job offer	45		E, F, S	no	n/a
PRB	both	to assist in the selection of reliable and conscientiousness employees	10-15	yes	E, S	no	-.26
PSI	applicants	to identify job applicants who tend to display CWB	15-60	yes	E, F, S	no	-.26
RR	applicants	to measure work-related conscientiousness and CWB as well as to predict job performance and CWB	15	yes	E, F, S	no	n/a
Stanton	applicants	to identify people acting counterproductively before hiring them	20-25	n/a	E	yes	n/a
VP (W6)	applicants	to identify suitable employees for non-exempt jobs	15-20	n/a	n/a	yes	n/a
Workkeys PA	applicants	to identify employees who tend to show problematic work behavior and who have particular working characteristics	10-15	yes	E	yes	-.27

Note. API = Applicant Potential Inventory, ARP = Applicant Risk Profiler, CBI = Counterproductive Behavior Index, ERI = Employee Reliability Inventory, ESQ = Employee Screening Questionnaire, HPI-R = Hogan Personality Inventory–Reliability Scale, ITEST = IntegriTEST, IBES = Inventar berufsbezogener und einstellungsbasierter Selbstbeurteilung [the Inventory of Job-Related Attitudes and Self-Evaluation], IPI-R = Inwald Personality Inventory–Revised, PRB = Personnel Reaction Blank, PSI = Personnel Selection Inventory, RR = Reid Report (the 29th edition), Stanton = Stanton Survey (the new edition), VP (W6) = ViewPoint (W6), Workkeys PA = Workkeys Performance Assessment.

"n/a" = information was not available, both = applicants and employees, E = English, F = French, S = Spanish, G = German, "various" = test is available in several languages.

[1] = All validity scores refer to self-reported data. All values are negatively represented because the correlation of integrity with CWB should be negative.

Inventory) and that the time specifications are based on descriptions by the test authors or publishers.

Aside from the duration of a test, the test format and language are essential criteria of test application. All integrity tests—with the exception of the Applicant Potential Inventory (due to its voice-guided application by telephone or via internet)—are available in a paper-and-pencil test format. Regarding the criterion of an available web-based format, an online version is offered for only ten tests. No web-based format is available for two tests, and no information about the test format is available for four tests. While web-based test formats have many advantages (e.g., the test can be carried out independent of time and location, and web-based test formats have been proven more effective in assuring objectivity than traditional test formats), data on the validity of web-based test formats are missing (Woods, Ahmed, Nikolaou, Costa & Anderson, 2020).

Due to the global application and reasons of fairness, it is essential to present a test in the appropriate language. The primary language of all tests is English, with the exception of the Inventory of Job-Related Attitudes and Self-Assessments, which is only offered in German.[5] About one-third of the available measures are published in a second (mostly Spanish) or third (mostly French) language (e.g., the Personnel Selection Inventory). In addition, tests often offer varieties of speech, such as British English or Vietnamese English. In some cases, numerous languages are available (e.g., the Hogan Personality Inventory–Reliability Scale is available in over 40 languages). Regarding the test application in different nations, it is important to gather data from these nations because their respective culture and attitude might differ (e.g., America and Japan have different working attitudes).

The criterion of the availability of a free manual was considered for two reasons: (a) In the Standards of Educational and Psychological Testing (Eignor, 2013), it is recommended that test material for test users be offered immediately when the test is published. (b) As previously mentioned, the lack of information has been widely bemoaned by researchers. A free manual is offered online for half of the tests (e.g., for the Stanton Survey–New Edition). In some cases, additional online information—such as item samples (e.g., for the Applicant Risk Profiler) or a user's guide (e.g., for the Counterproductive Behavior Index)—is freely available. Overall, half of the tests offer a free test manual.

[5] Test translations included in empirical studies were not considered.

The final criterion for test application is validity. As integrity tests per definition primarily refer to CWB, the criterion of validity was taken instead of other validity criteria (e.g., job performance). For validity, the predictive validity of CWB ranged from ρ = -.22 to -.66 (see Table 6). Although all reported validity scores refer to self-reported data, there is a wide range of validity values. In a comparison to meta-analytic findings (Ones et al., 1993; Van Iddekinge et al., 2012a), the overall mean observed correlation for CWB ranged from ρ = -.26 to -.33 (validity corrected for unreliability in the criterion: ρ = -.32 to -.47). Nevertheless, the maximum of these reported validity scores in this current overview is comparable to preceding meta-analytic findings. When analyzing measures with a very high validity score (>.45), these measures can be seen to include one or more dimensions of CWB as subscales (i.e., the subscale of drug avoidance or use (in most cases)). Only one measure—the IBES—does not refer to a CWB dimension. Upon closer examination of the test manual, no subscale or other criterion could be found to explain such a high validity score. Interestingly, this measure also contains both an overt as well as a personality-based part. Moreover, the overt part relates more significantly with CWB than does the personality-based part, which is in line with previous findings about the correlation of overt and personality-based measures with CWB (Ones et al., 1993; Van Iddekinge et al., 2012a).

In summary

With regard to their application, the aim of integrity tests not only consists in finding applicants with a tendency to display CWB but also in finding applicants with good job performance. This extension to predict job performance contrary to the straight definition of integrity tests to only predict CWB further weakens the construct of integrity. It is thus necessary for the aim of integrity tests to maintain a clear focus. To simplify test application, some test elements (e.g., duration, web-based format) should be considered when choosing an integrity test. Moreover, with regard to empirical data, gathering information about a used test via a free test manual is beneficial. With regard to validity, integrity tests show solid validity scores. Nevertheless, the criterion used for validity should be taken into consideration when evaluating the validity value of the test (e.g., self- vs. other-reported).

3.4 DISCUSSION

In the subsections of Chapter 3, integrity tests were analyzed both in overview and in detail to provide the state-of-the art of integrity tests. In general, integrity tests were found to be highly heterogeneous and overlapping with other constructs. This result has been confirmed by other researchers: Nicol and Paunonen (2002) found that two integrity tests did not measure the same construct. Moreover, Karren and Zacharias (2007) stated that "different integrity tests seem to focus on different attitudes or behaviors" (p. 223). In addition, scales that do not address the core concepts which integrity tests measure are used in research studies as a measure of integrity (Ones et al., 2012). As a result, Wanek (1996) advised caution when generalizing findings about overt integrity tests. It is therefore essential to consider integrity tests individually (Karren & Zacharias, 2007).

To support this idea of an individual focus on integrity tests, it may be helpful to develop a prototype integrity test which represents the main components of an integrity test. Moreover, a prototype can help to obtain a clear focus on integrity tests and can facilitate the interpretation and comparison of findings from different integrity tests. Such a prototype integrity test could be identified based on the analyses in this chapter.

In the following subsection, key findings about the status quo, quality, construction, and application of integrity tests are summarized. These findings are integrated into a prototype integrity test. Furthermore, a test that represents such a prototype—Göritz's (2014) 10-item integrity test[6]—is offered as a good example of a prototypical integrity test.

There is a predominance of overt integrity tests

The overview of integrity tests shows that the majority of identified integrity measures were overt tests (75%). For legal reasons, integrity tests had to be quickly developed in the early 90s because the polygraph and other physiological measurements were banned in personnel selection in the United States by the Employee Polygraph Protection Act (O'Bannon et al., 1989). Consequently, developers of integrity tests used variables that were in some way related to integrity in order to form integrity tests. While

[6] Anja Göritz worked as a professor at the Institute of Psychology at the University of Wuerzburg until 2011. There, the author got to know her and also received valuable advice from her with regard to the implementation of her own empirical studies included in this dissertation.

the majority of overt integrity tests was developed in the years after the Employee Polygraph Protection Act was passed, overt integrity tests are still developed today. One example of a more recent overt test is the 10-item integrity test, which was developed in 2014 (Göritz).

The quality of integrity tests varies

The development of integrity tests involves several challenges, and both theories and definitions of the integrity construct are missing in most cases. Moreover, in some cases, norm samples are also missing. Large and representative norm samples are essential with regard to the primary test purpose of the selection procedure. Free test information (e.g., validity data) and free test manuals are not available for all tests (Sackett, Lievens, Van Iddekinge, & Kuncel, 2017; Van Iddekinge et al., 2012a). Therefore, it can be difficult to evaluate a particular test's quality. These aspects should be kept in mind when choosing a measure to apply in research or in a selection procedure. In light of these challenges, the 10-item integrity test that reflects a narrow construct of integrity was carefully developed with regard to the prototypical items of integrity tests. The test also offers norm data on 915 participants. More than half of the norm sample is composed of working people (54% currently working, 24% in training or academic studies, and the rest, e.g., on parental leave or receiving a pension). Moreover, the test shows good internal consistency (α = .88; Göritz, 2014). Data and information are provided free of charge upon request.

The construct of integrity is confounded in integrity tests

The number and naming of subscales of integrity tests reveals a vast accumulation of every variable related to integrity. As a result, other constructs overlap with integrity and reflect a broad focus on this concept, which presents a threat to the construct validity of integrity tests. Although the integrity construct is multidimensional (e.g., J. Hogan & Brinkmeyer, 1997; Wanek et al., 2003) and relates strongly to several factors (e.g., conscientiousness, emotional stability, agreeableness), the use of other constructs or facets of constructs in integrity tests which are only weakly related to the construct of integrity falsifies construct validity. It is therefore essential for an integrity test used in empirical research to include fewer subscales and items. Moreover, both subscales and items should relate clearly and strongly to a narrow concept of integrity. The 10-item integrity test

complies with these requirements by using one scale with 10 items that address honest behavior toward others and honest actions in a private or work environment.

A clear aim of integrity tests is missing

Although integrity tests were originally developed for testing job applicants (ATP, 2010), some measures have other fields of application (e.g., employee development). In addition, although integrity tests were initially designed to predict CWB (Ones & Viswesvaran, 1998b), "job performance is the primary criterion of interest to most personnel selection specialists" (Sackett & Schmitt, 2012, p. 554). As a result, many tests aim to detect high-performing applicants rather than those with a tendency to display CWB. Meta-analyses have confirmed that integrity tests can predict job performance (Ones et al., 1993; Schmidt & Hunter, 1998; Van Iddekinge et al., 2012a). However, predicting job performance is not part of the classical definition of integrity tests, which does include the prediction of CWB (e.g., Ones & Viswesvaran, 1998b). The 10-item integrity test maintains this narrow focus on detecting CWB.

The application of integrity tests is not easy and quick

In order for integrity tests to be easy to use and apply, it is important to develop tests that are offered in a web-based format and can be completed quickly (Woods et al., 2020). Some current integrity tests offer no web-based format or last up to 60 minutes. In addition, in the current study, the average duration of integrity tests was found to be 20 minutes (*SD* = 10 minutes). In light of modern scenarios in both research (e.g., an online survey) and practice (e.g., an online personnel selection process), a short test duration of web-based tests is recommended by research studies (e.g., Marcus, Bošnjak, Lindner, Pilischenko, & Schütz, 2007). The 10-item integrity test provides both a web-based format and a short duration of 5 minutes or fewer.

In light of the goal of identifying a prototype integrity test and because of the ability of the 10-item integrity test to represent such a prototype, this test is further used in this dissertation for research on expanding the nomological network of integrity. Results about the relevant components of the nomological network of integrity are therefore more closely related to its fundamental construct and can be better generalized.

In addition, for the following research on the nomological network of integrity, it is particularly essential to use an overt integrity test like the 10-item integrity test for two reasons. First, overt integrity tests generally reveal higher correlations with validity criteria (Ones et al., 1993; Van Iddekinge et al., 2012a). Keeping in mind that a slight increase in validity can be essential for some test applications, an overt integrity test should be used for research on the validity of integrity. Second, overt integrity tests generally reveal lower correlations with personality traits. With regard to questions about personality trait correlates, overt tests serve as a more accurate yardstick. As a result, only strong correlations yield a significant result in the subsequent investigation on the nomological network of integrity.

CHAPTER 4 EXPANDING THE NOMOLOGICAL NETWORK OF INTEGRITY BY ADDING A PERSONALITY TRAIT: HONESTY-HUMILITY

In this chapter, the nomological network is expanded for the construct of integrity with a focus on personality traits. Conceptually, one of the most relevant personality traits for integrity is honesty-humility. Therefore, the relationship between both traits is investigated on the factor level as well as on the facet level of the honesty-humility construct. Moreover, integrity and honesty-humility strongly correlate with conscientiousness, which is one of the most relevant traits regarding the prediction of job-related behavior. Thus, the partial correlation between integrity and honesty-humility is investigated while controlling for conscientiousness. In addition, this chapter focuses on the incremental validity of an integrity test beyond honesty-humility and conscientiousness when predicting job performance.

4.1 THEORY AND HYPOTHESES

In recent decades, researchers have explored personality traits in terms of whether they correlate with integrity. Well-established personality factors such as the traits of the Five-Factor Model (Costa & McCrae, 1992) have been predominantly investigated (e.g., Marcus, Lee, & Ashton, 2007). Moreover, honesty-humility as a new sixth personality trait was introduced by the HEXACO model (Ashton & Lee, 2001). This model is in line with the Big Five and the Five-Factor Model with regard to the three factors extraversion, conscientiousness, and openness to experience. However, the personality trait models also differ in two aspects: (a) the two factors agreeableness and emotional stability represent rotated variants of those factors, and (b) the sixth factor honesty-humility is introduced.

Honesty-humility is defined as “the tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them” (Ashton & Lee, 2007, p. 156). People with a high level of honesty-humility are fair-minded and interact cooperatively with other individuals. They avoid manipulating others for personal gain and are not interested in wealth and luxuries. People with a low level of honesty-humility are dishonest and take advantage of other people or circumstances for

personal gain. They have a tendency to break rules for profit and are concerned with wealth and themselves.

As can be seen by the description of honesty-humility, the trait is conceptually related to CWB, which refers to intentional negative behavior that violates the legitimate interests of an organization that targets other people (e.g., coworkers) or the organization. Several studies have revealed negative moderate correlations of honesty-humility and CWB (Dalal, 2005; K. Lee, Ashton, & De Vries, 2005; K. Lee, Ashton, & Shin, 2005; O'Neill, Lewis, & Carswell, 2011). That is, people with a low level of honesty-humility are more likely to be involved in CWB. Moreover, honesty-humility is related to several CWB-relevant traits, such as Machiavellianism, narcissism, and psychopathy, which are referred to as the ''dark triad'' (K. Lee & Ashton, 2005; K. Lee et al., 2013).

Integrity and honesty-humility are both conceptually and statistically related: Honesty-humility significantly relates to overt integrity tests with a correlation about $r = .62$ and to personality-based integrity tests with a correlation about $r = .49$ (Marcus, Lee, & Ashton, 2007). This finding has been confirmed by various studies (e.g., K. Lee, Ashton, & De Vries, 2005; K. Lee, Ashton, Morrison, Cordery, & Dunlop, 2008). In addition, a meta-analysis found that integrity and honesty-humility strongly correlate with $\rho = .55$ and that no significant difference exists between corrected correlations for overt and personality-based integrity (Y. Lee, Berry, & Gonzalez-Mulé, 2019). Hence, some researchers even used the independent constructs integrity and honesty-humility interchangeably (e.g., A. De Vries, De Vries, & Born, 2011). The close relationship between both variables is supported by their connection to CWB (K. Lee, Ashton, & De Vries, 2005). Therefore, I suggest the following hypothesis:

H1a: Integrity is positively and significantly related to honesty-humility.

Like the Big Five factors, the personality trait of honesty-humility comprises a factor- and a facet level. Honesty-humility consists of four facets (K. Lee & Ashton, 2004): (a) *Sincerity* refers to the tendency to be authentic and act honestly in relationships with others, which means that people with low levels of sincerity may try to manipulate others for their own benefit. However, people with high levels of sincerity may be unwilling to take advantage of others. (b) *Fairness* describes the tendency to be fair-minded towards others

and to act honestly, which means that people with low levels of fairness tend to commit crimes, to cheat or steal. People with high levels of fairness tend to behave justly and lawfully. (c) *Greed avoidance* refers to being free of avarice, which means that people with low levels of greed avoidance may want to distinguish themselves from others and strive for money and fame. People with high levels of greed avoidance tend to be uninterested in luxury goods, status symbols, or privilege. (d) *Modesty* refers to being moderate and unassuming, which means that people with low levels of modesty may see themselves as being privileged and deserving special treatment. People with high levels of modesty see themselves as ordinary and often do not request special treatment.

Researchers have called for greater investigation into relations on the facet level of personality traits (e.g., Christiansen & Robie, 2011; Dudley, Orvis, Lebiecki, & Cortina, 2006; Rothstein & Goffin, 2006; Tett, Steele, & Beauregard, 2003). For example, narrow personality facets have predicted work outcomes, primarily job performance, at least as well as have broad personality factors, and these facets have explained additional variance (e.g., Dudley et al., 2006; Tett et al., 2003).

In addition to this statistical strength of narrow facets, methodological findings might also reveal better insights into the concept of integrity: In exploring the relationship between the facets of honesty-humility and integrity, suppression effects might be found (Chiaburu, Berry, & Gardner, & 2011). This effect means that a trait might not show a significant relationship to another trait on the factor level; however, its facets might exhibit significant relations in positive and negative directions.

Thus, with respect to the facet level of honesty-humility, I suggest the following hypothesis:

H1b: Integrity is positively related to facets of honesty-humility.

In addition to the conceptual relationship between integrity and honesty-humility, conscientiousness is another personality characteristic that is similar to both traits (K. Lee, Ashton, & De Vries, 2005). Conscientiousness, which is tantamount to being careful, responsible, organized, hardworking, and achievement-oriented, is part of the Five-Factor Model (Costa & McCrae, 1992) and—like honesty-humility—is also part of the HEXACO model (Ashton & Lee, 2001). The trait is a well-examined substantial correlate of integrity

(J. Hogan & Brinkmeyer, 1997; Marcus, Lee, & Ashton, 2007; Wanek et al., 2003) and honesty-humility (Marcus, Lee, & Ashton, 2007), both of which are moderately related to conscientiousness (Ashton et al., 2004). Moreover, conscientiousness is the best single predictor of integrity (Ones et al., 1993).

Although conscientiousness has been found to be the strongest correlate of CWB (Salgado, 2002), the HEXACO model—including honesty-humility—has substantially outperformed conscientiousness and the other four traits of the Five-Factor Model in predicting CWB and integrity (K. Lee, Ashton, & De Vries, 2005; Pletzer, Bentvelzen, Oostrom, & De Vries, 2019). Honesty-humility and integrity, in particular, are more related to CWB as the primary criterion of integrity tests than they are to conscientiousness (K. Lee, Ashton, & De Vries, 2005): "That is, the primary component of overt integrity tests is not the task-related conscience that characterizes the Conscientiousness factor but rather the moral conscience that characterizes the Honesty-Humility factor." (p. 192). With regard to performance at work, controlling for conscientiousness did not substantially diminish the correlation between integrity and job performance (Murphy & Lee, 1994a). Similar to this finding, the significant relationship between honesty-humility and CWB decreased considerably when integrity was partialled out, whereas the partial relationship between integrity and CWB remained on the same level when honesty-humility was partialled out (Laginess, 2016). Altogether, these findings imply that integrity consists of more components than conscientiousness and that it moreover relates more closely to honesty-humility. However, I suggest the following hypothesis:

H2: Partialling out conscientiousness reduces the positive correlation between integrity and honesty-humility non-significantly.

With regard to the primary criterion of CWB for integrity tests, integrity is additionally a substantial predictor of job performance (Ones et al., 1993; Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012a). Moreover, there is incremental validity for job performance when combining a measure of general mental ability with an integrity test as opposed to with a conscientiousness test (Schmidt & Hunter, 1998). In addition, integrity added validity in predicting behavioral integrity ratings over the HEXACO model and general mental ability (Marcus, Te Nijenhuis, et al., 2016). With regard to the prediction of CWB,

integrity also revealed to add validity beyond the HEXACO model, which includes honesty-humility (Marcus et al., 2013).

Regarding honesty-humility, this sixth personality factor also predicts CWB: Honesty-humility explains additional variance in CWB over conscientiousness (O'Neill et al., 2011). Moreover, honesty-humility has incremental validity for CWB over the Five-Factor Model including conscientiousness (Catano et al., 2018; K. Lee, Ashton, & De Vries, 2005; K. Lee, Ashton, & Shin, 2005; Pletzer et al., 2019). Conscientiousness and honesty-humility had incremental validity beyond the other HEXACO scales when predicting counterproductive academic behavior and even academic performance criteria (A. De Vries et al., 2011).

Research findings indicate conscientiousness as the best predictor of academic performance (e.g., Chamorro-Premuzic & Furnham, 2003; Noftle & Robins, 2007; Richardson & Abraham, 2009), which has been confirmed meta-analytically (O'Connor & Paunonen, 2007). Nevertheless, K. Lee et al. (2005) stated that the Five-Factor Model variables—including conscientiousness—cannot completely explain the high validity that integrity tests revealed when predicting supervisor ratings of job performance (e.g., Ones et al., 1993). Thus, I suggest the following hypothesis:

H3a: Integrity adds validity to job performance beyond conscientiousness and honesty-humility.

Next to honesty-humility, conscientiousness also consists of *facets* (K. Lee & Ashton, 2004): (a) *Organization* refers to favoring a structured order of equipment or process. (b) *Diligence* indicates being disciplined and hard-working. (c) *Perfectionism* is characterized by accuracy and being detail-oriented. (d) *Prudence* relates to acting with caution as opposed to acting on impulse.

To enhance predictive validity for different work criteria, the focus of research has changed from broad to narrow personality traits. The comparison of broad traits, which are more heterogenous, and narrow traits or facets, which are more homogeneous, is described in literature as the "bandwidth-fidelity" dilemma (Cronbach & Gleser, 1965; Ones & Viswesvaran, 1996). This means that the use of a broad trait (such as a trait of the Big Five or the HEXACO model) with a high bandwidth in the sense of high variability may entail a low fidelity. On the contrary, the use of a narrow trait (a facet of the Big Five or the HEXACO

model) leads to a low bandwidth, but a high fidelity outcome in the sense of accuracy or specificity about a criterion. Transferring the "bandwidth-fidelity" dilemma to work performance means that personality traits drive global performance while the facets drive specific performance.

Relying on broad traits may make a precise conceptual understanding of the relationship between personality and performance difficult to achieve, which has led some authors to conclude that narrower traits have more explanatory potential (Schneider, Hough, & Dunnette, 1996). Empirically, narrower facets have been shown to outperform broader trait domains (Pletzer, Oostrom, Bentvelzen, & De Vries, 2020). In consequence, when the goal is to maximize predictive validity for performance or to better understand the relations between personality and performance, it is preferable to investigate narrow traits as predictors (e.g., A. De Vries et al., 2011; Paunonen & Ashton, 2001).

Taking the facets of honesty-humility and conscientiousness constructs into account when predicting performance might explain more variance and thus may improve the prediction of performance criteria. Indeed, in the meta-analysis by Dudley et al. (2006), the conscientiousness facets had incremental validity for job performance over the conscientiousness factor. Furthermore, O'Neill et al. (2011) found that honesty-humility and its facet of fairness added validity to CWB beyond justice perceptions. Moreover, both the facets of conscientiousness and honesty-humility provided incremental validity when predicting academic performance criteria (A. De Vries et al., 2011). In comparison, the facets of both of these personality traits explained a higher percentage of the variance in academic performance criteria than did the factors of these two personality traits (A. De Vries et al., 2011). Hence, I posit the following hypothesis:

H3b: Integrity adds validity to job performance beyond the facets of conscientiousness and the facets of honesty-humility.

To provide a better overview, all hypotheses in this study are displayed in the following figure (see Figure 5):

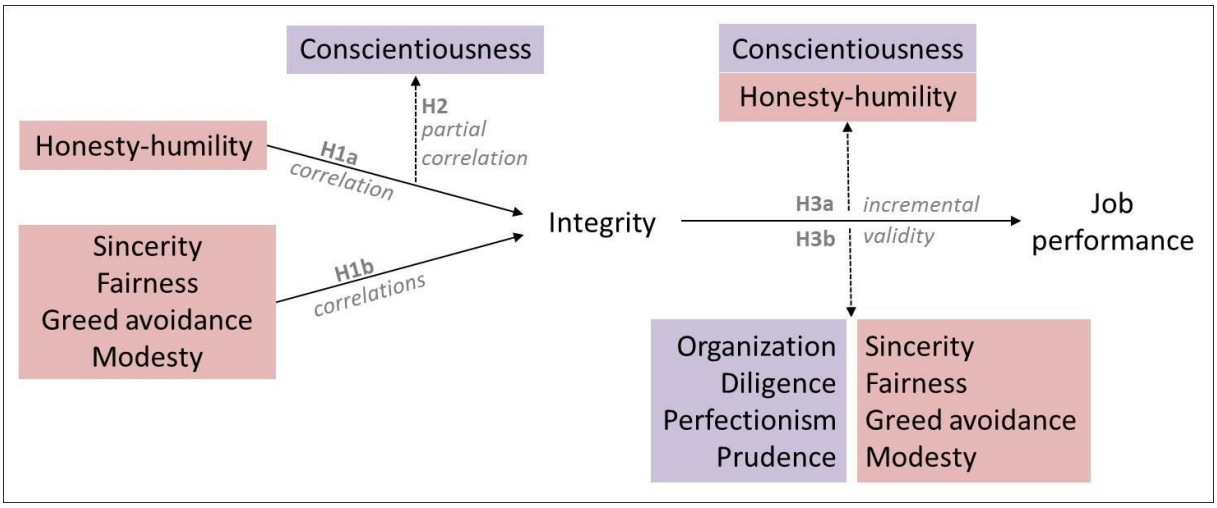

Figure 5. Overview of the hypotheses of Chapter 4 regarding the expansion of the nomological network of integrity with honesty-humility

4.2 METHODS

Data were collected online via a professional social networking site (www.xing.com), advertisement (http://wuewowas.de), and snowball sampling. Participants were rewarded by taking part in a drawing of vouchers. The software Unipark (http://www.unipark.de) was used for coding and saving the data.

Participants

A total of 269 participants completed the web-based survey. Participants were between 18 and 66 years, with an average age of 26 years (*SD* = 6.2). The majority of participants were women (71%), held a university entrance qualification (80%), and worked part-time (78%). Part-time employed participants frequently mentioned working as student assistants or research assistants at their university, sales assistants, or waiters/waitresses. Full-time employed participants indicated a variety of professions, including medical doctors, social-education workers, managers, and engineers. The average working time was 19 hours per week (*SD* = 15 hours), and one third of the participants claimed to work alone (32%).

Measures

Integrity. To measure integrity, the 10-item integrity test was used (Göritz, 2014). The test comprises ten overt items measuring attitudes toward dishonesty and fraud (e.g., "I can understand if somebody at work submits results that are not his/her own because of competitive pressure"; reversed). Participants answered on a 5-point rating scale ranging from 1 (*inaccurate*) to 5 (*accurate*). The internal consistency for the 10-item integrity test is α = .88 (Göritz, 2014).

Honesty-Humility. The personality trait of honesty-humility was measured using the German version of the HEXACO Personality Inventory-Revised (HEXACO-PI-R, Ashton & Lee, 2009). The inventory includes six major dimensions of personality (honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience). The honesty-humility scale comprises ten items related to four facets: (a) *sincerity* (e.g., "I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed"), (b) *fairness* (e.g., "If I knew that I could never get caught, I would be willing to steal a million dollars"; reversed), (c) *greed avoidance* (e.g., "Having a lot of money is not especially important to me"), and (d) *modesty* (e.g., "I think that I am entitled to more respect than the average person is"; reversed). Respondents answered on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal consistency for the honesty-humility subscale is α =.74 (Ashton & Lee, 2009).

Conscientiousness. The personality trait of conscientiousness was also measured using the German version of the HEXACO-PI-R (Ashton & Lee, 2009). The conscientiousness scale consists of ten items related to four facets: (a) *organization* (e.g., "I plan ahead and organize things, to avoid scrambling at the last minute"), (b) *diligence* (e.g., "I often push myself very hard when trying to achieve a goal"), (c) *perfectionism* (e.g., "I always try to be accurate in my work, even at the expense of time"), and (d) *prudence* (e.g., "I make decisions based on the feeling of the moment rather than on careful thought"; reversed). Respondents answered on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal consistency for the conscientiousness subscale is α = .76 (Ashton & Lee, 2009).

Job performance. Job performance is part of a German questionnaire aimed at determining performance-related work behavior (FELA-S; *Fragebogen zur Erfassung des leistungsbezogenen Arbeitsverhaltens*; Staufenbiel & Hartz, 2000). The questionnaire consists of five subscales, four of which related to facets of OCB (i.e., altruism,

conscientiousness, sportsmanship, and civic virtue). The job performance subscale was taken from a measure of in-role behavior of work (Williams & Anderson, 1991) and includes five items (e.g., "I perform my work tasks adequately"). The internal consistency for the job performance subscale is α = .91 (Staufenbiel & Hartz, 2000; Williams & Anderson, 1991).

Control variables. With regard to the predictors and criterion of the present study, the demographical variables of gender, age, and education were included as control variables.

4.3 RESULTS

Table 7 shows descriptive statistics and intercorrelations for integrity, honesty-humility, and conscientiousness on the factor- and facet level. Integrity and the factor honesty-humility correlate at $r = .45$, $p < .001$ (H1a). The analysis of the relations between integrity and the facets of honesty-humility reveal that all facets correlate at least at $p < .01$, ranging from $r = .16$ to .37 (H1b). Moreover, statistically controlling for conscientiousness did not appreciably reduce the correlation between integrity and honesty-humility to $r = .43$, $p < .001$ (H2). Using Fisher's *z*-transformation, the correlation of integrity with honesty-humility comparing with the correlation of integrity with honesty-humility while partialling out conscientiousness did not significantly differ (.45 vs. .43).

Table 7

Descriptive statistics and intercorrelations of integrity, honesty-humility, and conscientiousness on the factor- and facet level

	It.	*M*	*SD*	1	2	2a	2b	2c	2d	3	3a	3b	3c
1 10-item integrity test	10	4.19	0.62										
2 Honesty-humility	10	3.34	0.57	.45									
2a Sincerity	3	3.26	0.84	.25	.70								
2b Fairness	3	3.57	0.94	.37	.71	.26							
2c Greed avoidance	2	2.92	0.86	.16[a]	.57	.25	.15[b]						
2d Modesty	2	3.52	0.85	.34	.56	.18[a]	.21	.28					
3 Conscientiousness	10	3.62	0.56	.22	.14[b]	.20	.22	-.17[b]	-.01				
3a Organization	2	3.68	0.86	.16[a]	.13[b]	.19[a]	.18[a]	-.14[b]	.00	.74			
3b Diligence	2	3.80	0.74	.21	.11[c]	.11[c]	.21	-.13[b]	-.01	.70	.38		
3c Perfectionism	3	3.58	0.68	.14[b]	.06[c]	.15[b]	.14[b]	-.16[a]	-.09	.73	.31	.47	
3d Prudence	3	3.50	0.73	.17[a]	.13[b]	.14[b]	.15[b]	-.09	.06	.79	.55	.36	.36

Note. N = 269. It. = number of scale/subscale items. All correlations are $p < .001$ except where noted. [a] = $p < .01$. [b] = $p < .05$. [c] = non-significant

To test Hypotheses 3a and 3b, two block-wise hierarchical multiple regression analyses were conducted to determine the incremental validity of integrity beyond the factor- and facet level of honesty-humility and conscientiousness. In the first step, gender was added as a control variable because this variable has an effect on the predictors (A. De Vries et al., 2011; Marcus, Lee, & Ashton, 2007).[7] In the second and third step, conscientiousness and honesty-humility (their factor and their facets) were entered. In the fourth step, integrity was included.

[7] I also controlled for an effect of age and education, but no significant effect for either variable was found (for age: β = -.069, p = .265; for education: β = -.014, p = .812).

Table 8

Hierarchical regression analyses of job performance on the factor- and facet level of conscientiousness and honesty-humility as well as on integrity

	Job performance				Job performance		
Predictor (Factor level)	ΔR^2	β	*rw*	Predictor (Facet level)	ΔR^2	β	*rw*
Step 1	.03**			Step 1	.03**		
Gender		.05	5.6%	Gender		.06	4.6%
Step 2	.17***			Step 2	.20***		
Conscientiousness		.37***	57.4%	Organization		-.03	4.6%
				Diligence		.22***	24.8%
				Perfectionism		.10	12.3%
				Prudence		.18**	16.5%
Step 3	.01*			Step 3	.03**		
Honesty-humility		.00	4.6%	Fairness		.06	6.2%
				Greed avoidance		-.09	2.6%
				Modesty		.07	3.7%
				Sincerity		-.02	0.8%
Step 4	.05***			Step 4	.04***		
Integrity		.27***	32.4%	Integrity		.25***	23.9%
Total R^2_{adj}	.26***			Total R^2_{adj}	.27***		

Note. N = 269. *p < .05; **p < .01; ***p < .001. *rw* = relative weight (contribution to R^2)

Regarding the prediction of job performance on the factor level (H3a), integrity had significant incremental validity for job performance beyond the factor of conscientiousness and honesty-humility of ΔR^2 = .05, p < .001 (see Table 8, left panel). Overall, the traits conscientiousness and honesty-humility together with integrity explained variance in job performance at R^2_{adj} = .26, F(4,264) = 24.19, p < .001, BIC = 480.00.

The method of relative weight analysis was integrated to further explore the contribution of the personality variables when predicting job performance. By using this method, the relative importance of each predictor to the criterion could be determined while considering the unique contribution of each predictor and the contribution of each predictor in combination with the other predictors (J. W. Johnson, 2000). With regard to the relative weights, integrity accounted for 32.4% of the total variance in job performance, whereas conscientiousness accounted for 58.4% of the total variance in job performance.

Putting both traits together, integrity and conscientiousness accounted for 90.8% of the total variance in job performance.

Regarding the prediction of job performance on the facet level (H3b), integrity added validity in explaining job performance over the facets of conscientiousness and honesty-humility of $\Delta R^2 = .04$, $F(1, 258) = 14.47$, $p < .001$, (see Table 8, right panel). Overall, the facets of conscientiousness and honesty-humility together with integrity explained variance in job performance at $R^2_{adj} = .27$, $F(10, 258) = 11.11$, $p < .001$, BIC = 494.98.

Two of the four facets of conscientiousness significantly accounted for variance in job performance: the facet of diligence, $\beta = 0.22$, $t(257) = 3.49$, $p < .001$ and the facet of prudence $\beta = 0.18$, $t(257) = 2.76$, $p = .006$. In contrast, none of the four facets of honesty-humility significantly accounted for variance in job performance. Putting both predictors together, the facets of conscientiousness and honesty-humility accounted for 71.5% of the total variance in job performance. In comparison with the factor-level model, the facets of conscientiousness and honesty-humility together with integrity explained a slightly higher percentage of the total variance than did both broad traits of conscientiousness and honesty-humility together with integrity (95.4% vs. 94.4%).

With regard to the facet level, integrity accounted for 23.9% of the total variance in job performance. Comparable with this finding, the conscientiousness facet of diligence accounted for 24.8% of the total variance in job performance.

Comparing the factor- and facet models, the question remains which of the two models fit the data better. Considering the principle of parsimony (Forster, 2000), the facet model explained a slightly larger amount of variance in job performance ($R^2_{adj} = .26$ vs. $R^2_{adj} = .27$). It has to be considered that the facet model included more predictor variables at the expense of six degrees of freedom. Considering both model fit and parsimony, the factor model (BIC = 480.00) is superior to the facet model (BIC = 494.98), while a difference in BIC of 10 reveals a strong evidence (Raftery, 1995).

4.4 DISCUSSION

The key question of this chapter concerned the main constructs of the nomological network of integrity, particularly the relationship between integrity and honesty-humility.

This question refers to Hypotheses 1a and 1b. The data revealed that integrity and honesty-humility are closely related to each other on the factor- and facet level of honesty-humility. Moreover, Hypothesis 2 examined the relationship between integrity and honesty-humility while controlling for conscientiousness. The hypothesis that the relationship between integrity and honesty-humility does not decrease by partialling out conscientiousness was also confirmed by the results. Regarding the focus on the criterion, Hypotheses 3a and 3b investigated the incremental validity of integrity beyond the factor and facets of both conscientiousness and honesty-humility when predicting job performance. The findings revealed that integrity indeed adds incremental validity as a predictor of job performance beyond both the factor and facets of conscientiousness and honesty-humility. In the following section, all findings are discussed in detail and interpreted with regard to the nomological network of integrity.

Relation between integrity and honesty-humility

The present finding points to the conclusion that integrity and honesty-humility are two related concepts. In comparison with previous findings (K. Lee, Ashton, & De Vries, 2005; K. Lee et al., 2008; Y. Lee et al., 2019), the correlation between both personality variables in the current study (r = .45) is comparable to those found in previous studies. For example, Y. Lee et al. (2019) found a meta-analytic result of ρ = .55 for the relationship between integrity and honesty-humility. K. Lee et al. (2008) also found a correlation of r = .44 for integrity and honesty-humility, whereas K. Lee, Ashton, and De Vries (2005) found a slightly higher correlation of r = .53 for integrity and honesty-humility. Although K. Lee et al. (K. Lee, Ashton, & De Vries, 2005; K. Lee et al., 2008) used the same measure for honesty-humility as the present study, and although they both used an overt integrity test as did the present study, the studies were based on quite different samples: While the present study only referred to full- or part-time German employees, the other two studies of K. Lee et al. (K. Lee, Ashton, & De Vries, 2005; K. Lee et al., 2008) referred to a sample of Canadian university students. This difference may influence the value of integrity as a variable relating to a job position: Students may overestimate their integrity because they only have to imagine their behavior at work in contrast to employees, who must show integrity at work on a daily basis. Moreover, a difference in the degree of the relationship might be caused by an age effect of honesty-humility (Ashton & Lee, 2016): Honesty-humility showed an upward

trend of about one standard deviation between the ages of 18 and 60. In the current sample, participants' ages strongly varied from between 18 and 66 years, resulting in a rather low average age of 26 years (*SD* = 6.2).

The correlation between integrity and honesty-humility that was found in the present study is also comparable with the correlation that was found among overt integrity tests (Ones et al., 1993). In addition, researchers supported the congruence of both constructs (e.g., Laginess, 2016). This strong relationship between both traits indicates that honesty-humility has a high amount of identical conceptual connection to integrity, and that honesty-humility is comparable with the components of integrity tests. As a consequence, honesty-humility may become part of future integrity test developments.

In the present study, the correlation between integrity and honesty-humility was even notably higher than the correlation between integrity and conscientiousness, as expected when using an overt integrity test (e.g., Marcus, Lee, & Ashton, 2007). Nevertheless, in general, conscientiousness so far was found to be the strongest correlate of integrity (Berry, Sackett, & Wiemann, 2007). Therefore, it is not surprising that partialling conscientiousness out of the correlation between integrity and honesty-humility did not reduce the relationship between integrity and honesty-humility, in line with similar results for the relationship between integrity and job performance (Murphy & Lee, 1994a; Ones, 1993). In consequence, these findings further support the key role of honesty-humility for the construct of integrity (A. De Vries et al., 2011).

The relations of integrity and honesty-humility on the facet level revealed that every facet of honesty-humility correlates with integrity. Regarding the level of correlations, fairness and modesty showed to be more strongly connected to integrity than did the other facets of honesty-humility. This result is in line with the previous finding that fairness is the best predictor of academic counterproductive behavior (A. De Vries et al., 2011). With regard to the prediction of CWB, previous studies found the fairness facet of honesty-humility to be the most important predictor of delinquency (e.g., De Vries and Van Gelder, 2013; Pletzer et al., 2020). In addition, fairness was found to be the strongest correlate of Machiavellianism (Jonason & McCain, 2012).

Although these studies concentrated on counterproductive behavior and the personality trait of Machiavellianism, which is closely related to counterproductive behavior, the focus of the study at hand was on integrity itself. From a theoretical perspective, fairness

is part of several concepts of integrity, such as those of the AHRD (1999) and Barnard et al. (2008). This fact strengthens the present finding on the relationship between integrity and the fairness facet of honesty-humility. In conclusion, fairness seems to be related to integrity in its content: Fairness refers to acting honestly, not cheating or stealing, and obeying to law. Hence it is an important element of the construct of integrity.

Predicting job performance with integrity beyond conscientiousness and honesty-humility

The most important result regarding the prediction of job performance is that integrity added incremental validity beyond conscientiousness and honesty-humility. No prior study has examined this finding, although prior studies have concluded that the addition of honesty-humility explains incremental variance in the prediction of performance criteria (e.g., A. De Vries et al., 2011; M. K. Johnson, Rowatt, & Petrini, 2011). However, when integrity is included as a predictor, honesty-humility has no incremental validity for predicting task performance (Y. Lee et al., 2019). Both this result and the result of the current study support the ability of integrity to outperform other variables for predicting job performance.

Indeed, it is most notable that integrity predicts job performance beyond conscientiousness because the latter has been consistently shown to be the strongest predictor of job performance (e.g., Barrick & Mount, 1991; Salgado, 1997). Moreover, integrity explains additional variance in predicting job performance, although honesty-humility was found to be the best predictor of job performance beyond the Big Five factors (M. K. Johnson et al., 2011). Both variables of honesty-humility and conscientiousness, proved to add incremental validity to the prediction of job performance (A. De Vries et al., 2011). These findings are in line with the present finding that integrity predicts job performance beyond the factors of conscientiousness and honesty-humility. In contrast to these previous studies, the present study not only included conscientiousness and honesty-humility but also integrity in order to incrementally predict job performance.

The increase in validity found in the current study is comparable to that found in previous studies that have investigated the power of incremental validity that integrity tests have in predicting behavioral integrity ratings beyond the HEXACO model and general mental ability (ΔR^2 = .000 and .001) and in predicting CWB beyond the HEXACO model (ΔR^2_{IBES} = .034, .059) (Marcus et al., 2013; Marcus, Te Nijenhuis, et al., 2016). Depending on

which integrity test is used, some findings on gains in validity are higher (for the interity test based on the content of the California Psychological Inventory: $\Delta R^2_{CPI\text{-}Cp} = .090$ and for the Employee Integrity Inventory: $\Delta R^2_{EII} = .138$).

Although the empirical gain by adding integrity was moderate in the present study, it is important regarding applied situations, such as personnel selection. In terms of its theoretical focus, the present study provided another essential insight into the composition of the construct and nomological network of integrity because the finding clarifies the incremental predictive power of integrity with regard to job performance.

Previous studies have revealed that the facet level of conscientiousness (A. De Vries et al., 2011; Dudley et al., 2006) and of honesty-humility (A. De Vries et al., 2011) explained incremental validity in terms of predicting performance criteria. These findings are in line with the results of the present study, in which the facets of both traits—and the trait of integrity—added incremental validity to job performance. The present study thereby confirms previous findings and goes a step further with respect to predicting job performance.

With regard to the single facets of conscientiousness included in this survey, A. De Vries et al. (2011) found a similar pattern: The conscientiousness facet of diligence had the greatest significant gain ($\Delta R^2 = .18$, $p < .05$) in the study of A. De Vries et al. (2011). In comparison, the present findings also revealed that the conscientiousness facet of diligence produced the greatest significant gain ($\Delta R^2 = .22$, $p < .001$). Moreover, in the current study, the conscientiousness facet of prudence was found to have the second-highest significant gain; in contrast, prudence was not significant in the study of A. De Vries et al. (2011). This difference may be due to the fact that A. De Vries et al. (2011) used a sample of undergraduate students to predict academic performance, whereas the present study used a sample of employees to predict job performance. As a result, the facet of diligence, which refers to being disciplined and hard-working, might be relevant for both samples and prediction criteria, whereas prudence, which refers to being cautious and non-impulsive, might be more relevant with regard to employees in their job environment. Employees might more strongly avoid acting impulsively because of their older age and because of their greater life experience when compared with students.

With regard to the facets of honesty-humility, the present study revealed a different pattern in comparison with the findings of A. De Vries et al. (2011): No facet of

honesty-humility showed to be a significant predictor of job performance, whereas A. De Vries et al. (2011) found that the facets of greed avoidance and modesty significantly predicted academic performance. This difference in findings might again result from the difference of the investigated criterion, namely the difference in the used samples: Students are accustomed to having less money, luxury, and special treatment because of their status and limited income. These students concentrate more on earning a good degree in order to later find good employment. This is probably why A. De Vries et al. (2011) found that greed avoidance and modesty significantly predicted academic performance, in contrast to the non-significant finding in the present study, which examined employees and a job performance criterion.

Theoretical implications for the nomological network of integrity

Owing to the relationship between integrity and honesty-humility, the findings support the idea that the trait of honesty-humility—on the factor- and facet level—is substantially related to integrity. On a more detailed view, the facets of fairness and modesty revealed to be the most important bridge between integrity and honesty-humility. This finding is not surprising because both facets are conceptually related to integrity: A high level of fairness, which refers to having integrity and not cheating, stealing, being corrupt, or taking advantage of others, is closely linked to integrity, which refers to having no tendency to participate in any counterproductive behavior. Therefore, the facet of fairness is a basic aspect of the construct of integrity and should necessarily be considered in the nomological network of integrity.

The facet of modesty refers to being self-effacing and unassuming. As a consequence, people with a low level of modesty consider themselves superior and entitled to privileges that others do not have. Due to this description, low levels of modesty trigger a mindset that is common in people who behave counterproductively, whereas high levels of modesty prevent people from acting in a counterproductive manner. As a consequence, the facet of modesty is an important component of the nomological network of integrity.

The relationship between integrity and honesty-humility that was found in the present study is considerably stronger than the relationship between integrity and conscientiousness, as was expected by using an overt integrity measure. Moreover, partialling out conscientiousness did not reduce the correlation between both traits. These

two empirical findings are confirmed by the conceptual description of integrity and honesty-humility with respect to their constructs. These facts strengthen the notion that honesty-humility is an important part of the construct of integrity. In conclusion, honesty-humility should be considered a key concept of integrity.

Regarding the prediction of job performance in the present survey, the contribution of integrity was greater than that of honesty-humility, but lower than that of conscientiousness. On the facet level, only diligence and prudence as two facets of conscientiousness were shown to produce significant gain with regard to content. These two facets are the strongest correlates of integrity on the facet level of conscientiousness. A high level of the facet of diligence refers to being disciplined and hard-working, whereas a high level of the facet of prudence refers to acting carefully, not impulsively. In conclusion, both facets are empirically and conceptually related to integrity, which includes being and acting self-controlled and well-considered. Finally, diligence and prudence are identified as two important aspects in the nomological network of integrity.

Future research

There are additional traits that are similar to the concept of integrity. For example, self-control turned out to be a strong correlate of integrity (Bazzy et al., 2017). The relationship between both integrity and self-control should be considered in the context of additional personality correlates of integrity, e.g., honesty-humility. Comparing the single correlations and partial correlations of all traits might offer greater insight into the construct discussion about integrity.

Moreover, J. Hogan and Ones (1997) identified self-control as a facet of conscientiousness. In line with agreeableness and emotional stability, traits that include facets should be comprised, and their contribution to integrity should be investigated on the factor- and facet level. Hence, an analysis on the facet level of other traits in the Big Five model might add additional insights into the nomological network of integrity.

An additional interesting approach might be to compare the relations of the personality traits with regard to overt as well as personality-based integrity measures. In the past, both kinds of tests revealed different patterns concerning their relations to personality traits. For example, overt integrity tests demonstrated to be more strongly related to honesty-humility (K. Lee, Ahton, & De Vries, 2005; Marcus, Lee, & Ashton, 2007), whereas

personality-based integrity measures demonstrated to be more strongly correlated with conscientiousness (Marcus, Lee, & Ashton, 2007; Sackett & Wanek, 1996). As a consequence, the relationship between honesty-humility and the overt integrity test found in the present study might be lower when using personality-based integrity measures, whereas correlations with conscientiousness should improve. The question remains about the pattern of the facets relating to the different kinds of integrity test. It would be a progress to identify the pattern on the facet level of the traits by using a personality-based integrity test. In addition, the pattern of other familiar personality traits with both kinds of tests could reveal a clearer perspective on the context of the nomological network of integrity.

Strengths and limitations

While the present study contains some strengths, there are certain limitations to the content, mainly regarding the demographic and occupational characteristics of the sample, and at the methodological level.

One strength of the present study is the fairly large and diverse sample of employment, which may allow findings to be generalized more easily and to a broader range of jobs. The fact that the employees predominantly work part-time merits discussion because this might influence the variables of the study. Moreover, the sample consisted predominantly of women who were well-educated. In addition, compared with the general workforce, the participants were quite young, with a mean age of 26 years. Gender and educational level as well as age could have had an effect on integrity. However, other studies also examined relatively young samples: For example, Bourdage, Lee, Lee, and Shin (2012) referred to a sample with an average age of 29 years, and K. Lee, Ashton, and De Vries (2005) used a sample with an average age of both 26 and 21 years.

A further limitation of the present study is its use of self-reports given by the same person for all variables. Hence, the relations found might have been inflated due to common method variance and single source bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Nevertheless, a meta-analysis revealed that supervisor ratings of job performance highly correlate with self-ratings of job performance (e.g., Heneman, 1986). Moreover, K. Lee and Ashton (2013) found that self- and observer-reported data of the HEXACO are comparable, while incurring a small bias in both sources. In addition, Ashton and Lee (2010)

demonstrated that self-reports measure some facets of honesty-humility more accurately than do other-reports. This finding might result from the fact that the trait of honesty-humility and its facets are difficult to observe for any person.

Integrity tests showed to be prone to social desirability (Alliger & Dwight, 2000). Regarding self-reported data, the effect of social desirability might have occurred. Although there was no control for social desirability in this study, the influence of socially desirable answering was reduced by the setting: In an anonymous web-based setting—as in the present case—social desirability occurs less than in a non-anonymous setting (Crutzen & Göritz, 2011). Nevertheless, if there was an influence of social desirability, it would likely apply to all variables in the study because all of them reflect a desirable trait or behavior. All values may thereby increase, but the effects regarding the correlations and the hierarchical linear regressions should remain the same.

Conclusion

Personality traits have been proven to be the most meaningful antecedents of job-related behavior (e.g., Salgado, 1997). This study focused on personality traits and extended the nomological network of integrity by investigating and comparing nearby personality traits on their factor- and facet level. Both traits of honesty-humility and conscientiousness appear to be significant correlates of integrity, whereas integrity appears to be a substantial predictor of job performance beyond the factor and facets of both honesty-humility and conscientiousness. This study also demonstrated that integrity is strongly related to honesty-humility beyond conscientiousness and might henceforth be more intensely considered in the prediction of job performance.

Implications for the construct and nomological network of integrity can be made: The honesty-humility factor—especially its facets of fairness and modesty—contributes as a key concept to the construct of integrity. In contrast to the factor of conscientiousness, its facets of diligence and prudence are more strongly related to integrity and therefore represent two necessary components in the construct of integrity. Finally, the present findings should encourage researchers to further investigate the nomological network while concentrating on additional personality antecedents of integrity on both the factor- and facet level.

CHAPTER 5 EXPANDING THE NOMOLOGICAL NETWORK OF INTEGRITY BY ADDING A BEHAVIORAL CHARACTERISTIC: OCB

In this chapter, the nomological network of integrity is expanded to include the construct of integrity with a focus on organizational citizenship behavior (OCB). OCB means individual behaviors that are beneficial to the organization or its members but are not directly recognized by any formal reward system. Given that the construct is multidimensional, it consists of five facets described by Organ (1988): (a) Altruism refers to behaviors directed toward helping or cooperating with coworkers. (b) Conscientiousness describes employees' behaviors to accept and comply with organizational rules. (c) Sportsmanship refers to a willingness to tolerate and not to complain about minor workplace inconveniences. (d) Courtesy reflects behaviors that help to prevent problems in the workplace. (e) Civic virtue refers to responsible and constructive political involvement in organizations. OCB is conceptually and empirically related to both CWB and conscientiousness—two variables that are closely related to integrity. While integrity research has thus far predominantly focused on CWB and conscientiousness, the link between integrity and OCB, by contrast, has been neglected for a long time. Therefore, in the first step, the relationship between integrity and OCB was investigated on the factor as well as facet level of OCB. Moreover, the relationship was examined in terms of the influence of conscientiousness. In the second step, the ability of integrity to predict OCB was explored. The incremental validity of OCB caused by integrity beyond conscientiousness and honesty-humility—both personality traits were examined on the factor as well facet level—was thus tested.

5.1 THEORY AND HYPOTHESES

Integrity is a well-known personality construct that correlates with job performance (Ones et al., 1993; Schmidt & Hunter, 1998). Job performance can be conceived of as a sum of different individual behaviors that occur over a standard period of time (Motowidlo, 2003). There are three main dimensions of job performance (Borman & Motowidlo, 1993; Rotundo & Sackett, 2002): First, task performance is defined as "the proficiency with which incumbents perform activities that are formally recognized as part of their jobs" (Borman &

Motowidlo, 1993, p. 73). Second, the contextual performance of CWB refers to intentional negative behavior that hurts the legitimate interests of an organization and has the aim of harming members of the organization or the organization itself (Nerdinger et al., 2008). Third, the contextual performance of OCB refers to employees' behavior of putting in extra effort beyond the formal job requirements of the employment contract (Organ, 1988, 1997). Therefore, OCB is not rewarded by any formal reward system within the organization (Organ, 1988). However, an employee benefits from OCB indirectly, for example, by receiving a good performance assessment (Organ, 1997).

Job and task performance as well as CWB have adequately empirically proven to relate strongly to integrity, whereas comparatively less research has been conducted on the relationship between integrity and OCB (Van Iddekinge et al., 2012a). With regard to job and task performance, the correlation of self-reported integrity with supervisors' ratings of job performance among over 18,000 working participants across 90 independent samples was $\rho = .33$ (Ones et al., 1993). Moreover, Schmidt and Hunter (1998) found that integrity adds remarkable validity to the prediction of job performance in addition to taking a general mental ability test. With regard to the contextual performance of CWB, meta-analyses of relevant studies have found integrity to be a strong personality correlate of CWB (Berry, Sackett, & Wiemann, 2007; Ones et al., 1993; Van Iddekinge et al., 2012a). Additionally, the primary validity criterion of integrity tests predicts CWB (Ones et al., 1993; Van Iddekinge et al., 2012a). A meta-analysis found integrity tests to predict CWB with a mean validity of $\rho = .32$ (Van Iddekinge et al., 2012a). Furthermore, integrity tests have generally been found to predict different forms of CWB approximately equally well. In general, integrity tests have been developed to predict CWB based on a criterion-oriented approach rather than on a construct-oriented approach, which means that test items are a-theoretically chosen due to their relationship with the criterion of CWB (Berry, Sackett, & Wiemann, 2007).

In contrast to these studies and meta-analyses on integrity and CWB and on job and task performance, only few studies has empirically addressed the relationship between integrity and OCB. For example, Casillas, Robbins, McKinniss, Postlethwaite, and Oh (2009) found a correlation of $r = .70$ between integrity and OCB. Although this correlation is substantial, it should be noted that the study only investigated "narrow facets of an integrity test" (p.119), and therefore had a restricted focus on integrity. This restriction leads one to overestimate the relationship because integrity measures are defined to have a general

focus on CWB in order to predict CWB (Ones & Viswesvaran, 1998b): (a) The first subscale of their integrity test—"general work attitudes"—referred to job performance and OCB instead of referring to CWB. Thus, both variables partially referred to the same construct. (b) The content of the second subscale of their integrity test—"risk reduction"—referred more to OCB than to CWB. Consequentially, the question of whether integrity predicts OCB has to be clearly empirically answered.

Irrespective of an empirical link between both variables, the concepts of OCB and integrity are similar: OCB is a positive behavior that has benefits for the organization (i.e., participating in organizational events or making a suggestion for improving a particular work procedure) and supports coworkers (i.e., helping coworkers with a work task) (Organ, 1988, 1997). This characteristic in line with the concept of integrity, which is also a positive behavior that benefits the organization (i.e., being honest with working time or not damaging organizational property) or members of the organization (i.e., not attacking coworkers). Moreover, both concepts relate similarly to the same personality characteristics (e.g., conscientiousness) and to the same organizational behavior (e.g., CWB).

With regard to contextual performance, the strong relationship between CWB and OCB has to be considered in the network of integrity. OCB is theoretically and empirically related to CWB in two ways: (a) CWB is theoretically linked to OCB as contextual performance in Borman and Motowidlo's (1993) performance model. The kind of contextual performance means not to consider to working tasks defined by job contracts, but consider to working tasks beyond the requirements in job contracts. Contextual performance—in contrast to task performance—is influenced more strongly by personality factors than by cognitive abilities (Borman & Motowidlo, 1993; Motowidlo, Borman, & Schmitt, 1997). (b) CWB is also empirically related to an OCB of up to ρ = -.32 (Dalal, 2005) to -.39 (Sackett, Berry, Wiemann, & Laczo, 2006). Moreover, CWB and OCB are both predicted by similar non-personality factors (e.g., job satisfaction) and by similar personality factors (e.g., conscientiousness; Bowling, 2010).

As a result of the strong conceptual and empirical relationship between CWB and OCB, researchers have called for exploring the relationship between integrity and OCB (e.g., Hertel et al., 2000). Moreover, as integrity predicts CWB, which is an important correlate of OCB, integrity can also be assumed to predict OCB. In their meta-analysis, Van Iddekinge et al. (2012a) reported a corrected criterion-related validity of integrity for predicting OCB

(referred to as contextual performance) of ρ = .14. These relations lead to the following hypothesis:

H1a: Integrity is positively and significantly related to OCB.

Based on the relationship between both variables, the next step is to conduct a detailed examination of the relationship between both concepts. As integrity is a multifaceted construct (Berry, Sackett, & Wiemann, 2007), there has not been a common understanding or definition of its facets. In contrast, OCB is a complex construct that consists of five facets (Organ, 1988). As explicated in Chapter 4, there are important reasons to investigate variables on their facet level. Facets can explain additional variance (e.g., Dudley et al., 2006; Tett et al., 2003), provide clearer insights into conceptual relationships with the criterion (Hastings & O'Neill, 2009), and be used to find methodological restrictions (e.g., suppression effects) (Chiaburu et al., 2011). Facets are generally defined more specifically so that their relationships with variables are easier to predict and understand (Schneider et al., 1996). As OCB and integrity are theoretically and empirically similar (Sackett, Berry, Wiemann, & Laczo, 2006; Dalal, 2005; Borman & Motowidlo, 1993), all facets of OCB can be assumed to correlate positively with integrity. Thus, with respect to the facet level of OCB, I suggest the following hypothesis:

H1b: Integrity is positively related to facets of OCB.

An important personality trait that correlates moderately with integrity is conscientiousness (J. Hogan & Brinkmeyer, 1997; Marcus, Lee, & Ashton, 2007; Wanek et al., 2003), which has proven a consistent positive correlate of integrity (Chiaburu et al., 2011; Murphy & Lee, 1994b). In addition, there are also similarities in the items and subscales of the measures for both constructs (J. Hogan & Hogan, 1989). Moreover, Berry, Sackett, and Wiemann (2007) stated that integrity seems to be a hierarchical construct that refers to as an overall conscientiousness factor. Similar to integrity, conscientiousness is also positively related to job performance (Barrick & Mount, 1991; Schmidt, Shaffer, & Oh, 2008; Tett et al., 1991), and negatively related to CWB (Berry, Ones, & Sackett, 2007; Sackett & DeVore, 2001).

With regard to OCB, there is evidence suggesting that personality traits are more relevant to contextual performance such as OCB than they are to task performance (e.g., Borman & Motowidlo, 1993; Motowidlo et al., 1997; Motowidlo & Van Scotter, 1994). Different studies have found conscientiousness to be a well-examined predictor of OCB (Hoon & Tan, 2008; Organ, 1994; Organ & Ryan, 1995; Podsakoff et al., 2000). Meta-analyses have shown that conscientiousness is a better predictor of OCB than are other factors of the Five-Factor Model, such as agreeableness (e.g., Ilies, Fulmer, Spitzmuller, & Johnson, 2009). These findings imply that integrity and OCB consist of more components than conscientiousness and that integrity is moreover closely related to OCB. However, I suggest the following hypothesis:

H2: Partialling out conscientiousness reduces the positive correlation between integrity and OCB non-significantly.

In light of the validity of integrity tests, researchers have called for an investigation into whether integrity predicts OCB (e.g., Hertel et al., 2000). Surprisingly, relatively little research has been performed on this topic compared with the vast field of research on integrity and CWB, but a few empirical studies that deal with this prediction exist. While some studies have found non-significant predictive power of integrity for OCB (e.g., Y. Lee et al., 2019), other studies, in contrast, have found vast predictive power of integrity for OCB (e.g., Casillas et al., 2009).

Casillas et al. (2009) found that OCB has a predictive power of integrity that explains 49% of variance in OCB. This value is an impressive effect size, although the association was overestimated due to the reasons explained above (the integrity measure had no clear focus on the construct and, instead referred strongly to the OCB construct). Moreover, with regard to test validity, the influence of corresponding variables, which are related to the predicted variable, should be controlled for. To do so, two variables are identified to control for: conscientiousness and honesty-humility.

A meaningful and stabile characteristic for predicting OCB is conscientiousness (Berry, Ones, & Sackett, 2007; Ilies et al., 2009). Several studies have demonstrated conscientiousness to be a well-examined predictor of OCB (Hoon & Tan, 2008; Organ, 1994; Organ & Ryan, 1995; Podsakoff et al., 2000). Meta-analyses have revealed conscientiousness

to be a better predictor of OCB than other factors of the Five-Factor Model, such as agreeableness (e.g., Ilies et al., 2009; Organ & Ryan, 1995). Indeed, conscientiousness is the Five-Factor Model personality trait that best predicts OCB (Borman, Penner, Allen, & Motowidlo, 2001; Bowling, 2010; O'Brien & Allen, 2007).

In light of the theoretical and empirical relationship between honesty-humility and CWB, Chiaburu et al. (2011) have suggested that honesty-humility may also be "meaningfully related to citizenship" (p. 13). Only a few studies on this relationship exist: For example, Pedooem (2007) found honesty-humility to be significantly related to (parts of) OCB (r = .50); however, the measure of OCB she used focused on only two facets of OCB: interpersonal helping and rule following. To give another example, Bourdage et al. (2012) found honesty-humility to predict a certain kind of OCB: Individuals low on honesty-humility were motivated to engage in OCB due to impression management (standardized path coefficient = -.30). Honesty-humility was not related to OCB that was motivated by other motives (e.g., prosocial values, organizational concerns). Bourdage et al. (2012) have suggested that the three investigated OCB motives may not be exhaustive. Moreover, OCB referring to individuals (OCB-I) and to the organization (OCB-O) was not correlated with honesty-humility.

Like integrity and conscientiousness, honesty-humility also shows the same relationships to job performance and CWB: Honesty-humility was found to be negatively and moderately related to CWB (Dalal, 2005; K. Lee, Ashton, & De Vries, 2005; O'Neill et al., 2011). Furthermore, honesty-humility also added incremental validity to CWB over the Five-Factor model by providing considerable practical significance (Catano et al., 2018; K. Lee, Ashton, & De Vries, 2005). With regard to the prediction of job performance, honesty-humility also added incremental validity to job performance over the Big Five factors (Johanson, Rowatt, & Petrini, 2011) and accounted for variability over the Big Five factors (Ashton & Lee, 2008).

These findings imply that integrity, conscientiousness, and honesty-humility are characteristics that relate closely to OCB and that conscientiousness best predicts OCB. In addition, integrity was also found to predict OCB (Van Iddekinge et al., 2012a), but nothing is known about its incremental validity beyond other related traits, such as conscientiousness. Taken together, these findings on the relationship between and prediction of the mentioned variables imply that integrity appears to have predictive power with regard to OCB beyond

conscientiousness and honesty-humility. Therefore, these findings lead to the following hypothesis:

H3a: Integrity adds validity to OCB beyond conscientiousness and honesty-humility.

Like OCB, conscientiousness and honesty-humility consist of facets (K. Lee & Ashton, 2004). Narrow personality facets have predicted work outcomes at least as well as broad personality factors and have explained additional variance (Dudley et al., 2006; Hastings & O'Neill, 2009; Tett et al., 2003). Moreover, facets outperform factors in the prediction of job performance (Rothstein & Goffin, 2006) and of CWB (Pletzer et al., 2020). However, the facet level helps to identify the narrow traits that are in charge of predicting OCB (Thielmann, Spadaro, & Balliet, 2020).

Studies on the relationship between honesty-humility and OCB (e.g., Bourdage et al., 2012; Pedooem, 2007) have not taken the facets of honesty-humility and OCB into account. Therefore, with respect to the facet level of conscientiousness (organization, diligence, perfectionism, prudence) and honesty-humility (sincerity, fairness, greed avoidance, modesty), I suggest the following hypothesis:

H3b: Integrity adds validity to OCB beyond the facets of conscientiousness and the facets of honesty-humility.

To provide a better overview, all hypotheses in this study are displayed in the following figure (see Figure 6):

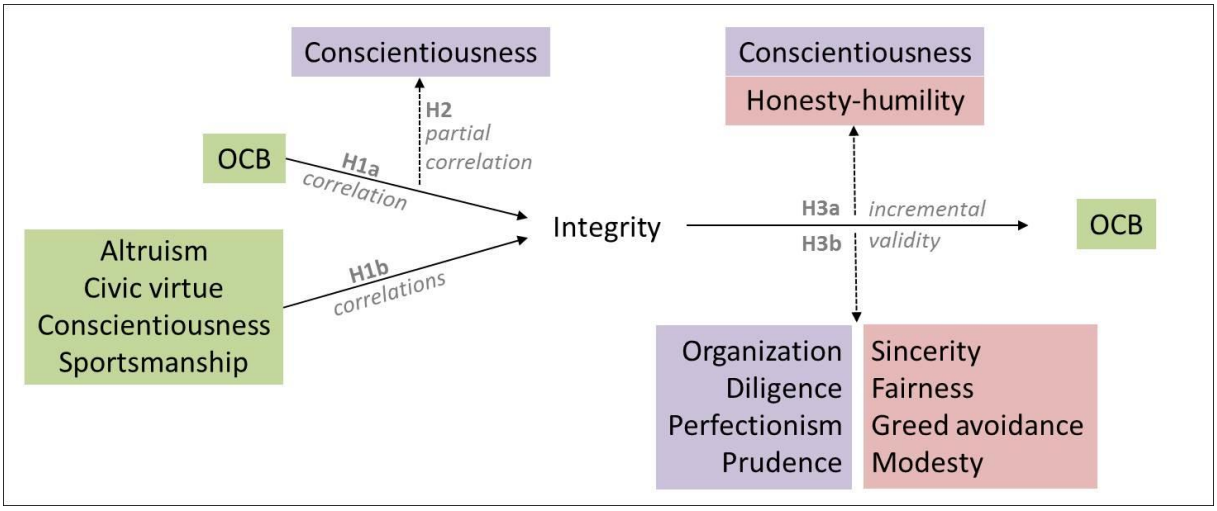

Figure 6. Overview of the hypotheses of Chapter 5 regarding the expansion of the nomological network of integrity to include OCB

5.2 METHODS

As mentioned in Chapter 4, data were collected online via a professional social networking site, advertisements, and snowball sampling. Participants were rewarded by taking part in a drawing of vouchers, and the software Unipark was used for coding and saving the data.

Participants

A total of 269 participants completed the web-based survey. As OCB refers to work performance beyond the job description of a working contract, participants without a working contract were excluded from data analysis ($n = 61$). As a consequence, these participants were also excluded from further analyses. The final sample consisted of 208 participants between 18 and 57 years, with an average age of 26 years ($SD = 5.8$ years). More than half of the participants were women (68%), held a university entrance qualification (79%), and worked part-time (74%). The average working time was 21 hours per week ($SD = 15.3$ hours). One-quarter of the participants claimed to work alone (25%), but did in fact have colleagues (e.g., research assistants).

Measures

Integrity. To measure integrity, the 10-item overt integrity test was used (Göritz, 2014). The test comprises ten overt items measuring attitudes toward dishonesty and fraud (e.g., "I can understand if somebody at work submits results that are not his/her own because of the pressure of competition"; reversed). An overt integrity test (instead of a personality-based test that is grounded on personality factors) was chosen because of its obvious relationship to CWB. This kind of integrity test refers to external criteria such as counterproductive attitudes, admissions of past dishonest behavior, or rationalizations of CWB and its precursors. Therefore, this kind of test is closer to the concept of CWB and thus also closer to the concept of OCB. Participants answered on a scale ranging from 1 (*inaccurate*) to 5 (*accurate*). The internal consistency for the 10-item integrity test is α = .88 (Göritz, 2014).

OCB. A German questionnaire based on Organ's OCB taxonomy (1988), was used to determine OCB (FELA-S; *Fragebogen zur Erfassung des leistungsbezogenen Arbeitsverhaltens*; Staufenbiel & Hartz, 2000). The questionnaire measures four of the five OCB facets, each with five items: altruism (e.g., "I support coworkers when they are overloaded with work"), conscientiousness (e.g., "I always get to work on time"), sportsmanship (e.g., "I spend a lot of time complaining about trivial things"; reversed), and civic virtue (e.g., "I keep up to date about new organizational developments"). The fifth facet of OCB, which is courtesy, was not replicated as a factor of its own and therefore not integrated as a subscale in the questionnaire. Participants rated the extent to which they engage in OCB on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). For the four facets of OCB, internal consistencies of α =.87 for altruism, α = .76 for conscientiousness, α =.86 for sportsmanship, α =.87 for civic virtue, and α =.92 for the entire OCB questionnaire were reported (Staufenbiel & Hartz, 2000).

Conscientiousness. The personality trait of conscientiousness was measured via the German version of the HEXACO-PI-R (Ashton & Lee, 2009). The conscientiousness scale consists of ten items that relate to four facets: (a) *organization* (e.g., "I plan ahead and organize things to avoid rushing at the last minute"), (b) *diligence* (e.g., "I often push myself very hard when trying to achieve a goal"), (c) *perfectionism* (e.g., "I always try to be accurate in my work, even at the expense of time"), and (d) *prudence* (e.g., "I make decisions based on how I feel in the moment rather than on careful thought"; reversed). Respondents

answered on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal consistency for the conscientiousness subscale is α = .76 (Ashton & Lee, 2009).

Honesty-Humility. The personality trait of honesty-humility was also measured using the German version of the HEXACO Personality Inventory-Revised (HEXACO-PI-R, Ashton & Lee, 2009). The honesty-humility scale comprises ten items related to four facets: (a) *sincerity* (e.g., "I wouldn't use flattery to get a raise or promotion at work, even if I thought it would work"), (b) *fairness* (e.g., "If I knew that I could never get caught, I would be willing to steal a million dollars"; reversed), (c) *greed avoidance* (e.g., "Having a lot of money is not particularly important to me"), and (d) *modesty* (e.g., "I think that I am entitled to more respect than the average person"; reversed). Respondents answered on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal consistency for the honesty-humility subscale is α =.74 (Ashton & Lee, 2009).

Control variables. Regarding the predictors and criterion of the present study, the demographical variables of gender, age, and education were included as control variables. Moreover, to control for variables that possibly affected OCB, participants' weekly working time (based on their formal employment contract), their actual weekly working time, the type of remuneration (i.e., fixed hourly wage, pay related to individual performance, pay related to group performance, pay related to both individual and group performance), and the number of coworkers were recorded.

5.3 RESULTS

Table 9 presents descriptive statistics and intercorrelations among integrity and OCB as well as conscientiousness on the factor- and facet level. Integrity and OCB correlated at r = .37, p < .001 (H1a). This finding confirmed Hypothesis 1a. The analysis of the relationships between integrity and the facets of OCB revealed that three of four facets correlated at p < .001 and ranged from r = .25 to .36. The OCB facet of civic virtue did not significantly correlate with integrity, p = .331 (H1b). These findings did not confirm Hypothesis 1b on the whole.

The next hypothesis H2 stated that partialling out conscientiousness would reduce the positive correlation between integrity and OCB non-significantly. Results revealed that

statistically controlling for conscientiousness did not appreciably reduce the correlation between integrity and OCB to $r = .32$, $p < .001$ and thereby supported H2. When comparing the correlation of integrity with OCB with the correlation of integrity with OCB while partialling out conscientiousness using Fisher's *z*-transformation ($r = .37$ vs. $r = .32$), both correlations did not significantly differ.

To test Hypotheses 3a and 3b, two block-wise hierarchical multiple regression analyses were conducted to determine the incremental validity of integrity beyond the factor and facets of both conscientiousness and honesty-humility (see Table 10).[8] In the first step, conscientiousness (factor and facets) was entered. In the second step, honesty-humility (factor and facets) was entered. In the third step, integrity was included.

Regarding the prediction of OCB on the factor level (H3a), integrity had significant incremental validity for OCB beyond the factor of conscientiousness and honesty-humility $\Delta R^2 = .06$, $p < .001$ (see Table 10, left panel). Overall, the traits honesty-humility and conscientiousness together with integrity explained variance in OCB at $R^2_{adj} = .24$, $F(3,204) = 22.42$, $p < .001$, BIC = 964.30.

As in Chapter 4, the method of relative weight analysis was integrated to further explore the contribution of the variables when predicting OCB. The relative importance of each predictor to the criterion could be determined while considering the unique contribution of each predictor and the contribution of each predictor in combination with the other predictors (J. W. Johnson, 2000). With regard to the relative weights, integrity accounted for 37.9% of the total variance in OCB, whereas conscientiousness accounted for 50.6% of the total variance in OCB. Putting both traits together, integrity and conscientiousness accounted for 88.5% of the total variance in OCB.

Regarding the prediction of OCB on the facet level (H3b), integrity added validity in explaining OCB over the facets of conscientiousness and honesty-humility of $\Delta R^2 = .04$, $p < .001$, (see Table 10, right panel). Overall, the facets of conscientiousness and honesty-humility together with integrity explained variance in OCB at $R^2_{adj} = .33$, $F(9, 198) = 10{,}82$, $p < .001$, BIC = 949.30.

[8] I controlled for an effect of gender, age, and education, but no significant effect for either variable was found (for gender: $\beta = -.015$, $p = .823$; for age: $\beta = .067$, $p = .299$; for education: $\beta = .032$, $p = .600$).

Table 9

Descriptive statistics and intercorrelations of integrity, OCB, and conscientiousness on the factor- and facet level

	It.	*M*	*SD*	1	2	2a	2b	2c	2d	3	3a	3b	3c	3d	4	4a	4b	4c
1 10-item integrity test	10	4.22	0.61															
2 OCB	20	5.52	0.53	.37														
2a Altruism	5	5.43	0.76	.24	.70													
2b Civic virtue	5	4.87	1.07	.07 [c]	.71	.45												
2c Conscientiousness	5	5.77	0.83	.32	.62	.23	.17 [b]											
2d Sportsmanship	5	5.44	0.83	.36	.53	.16 [b]	.03 [c]	.23										
3 Conscientiousness	10	3.64	0.52	.20	.39	.08 [c]	.29	.32	.29									
3a Organization	2	3.69	0.85	.14 [b]	.28	.10 [c]	.19 [a]	.23	.17 [b]	.75								
3b Diligence	2	3.81	0.74	.22 [a]	.41	.15 [b]	.26	.31	.32	.70	.39							
3c Perfectionism	3	3.59	0.65	.09 [c]	.24	.03 [c]	.22	.23	.10 [c]	.69	.27	.44						
3d Prudence	3	3.55	0.70	.13 [c]	.25	-.01 [c]	.17 [b]	.18 [a]	.27	.76	.54	.31	.26					
4 Honesty-humility	10	3.32	0.57	.43	.25	.19	.05 [c]	.20 [a]	.23	.18 [a]	.16 [b]	.13 [c]	.06 [c]	.18 [a]				
4a Sincerity	3	3.21	0.82	.24	.12 [c]	.07 [c]	.08 [c]	.09 [c]	.07 [c]	.23	.19 [a]	.14 [b]	.19 [a]	.15 [b]	.70			
4b Fairness	3	3.57	0.94	.37	.39	.25	.25	.28	.21 [a]	.25	.21 [a]	.22 [a]	.13 [c]	.19 [a]	.72	.29		
4c Greed avoidance	2	2.88	0.87	.18 [b]	-.08 [c]	.00 [c]	-.17 [b]	-.07 [c]	.06 [c]	-.10 [c]	-.08 [c]	-.10 [c]	-.14 [b]	.00 [c]	.57	.20 [a]	.17 [b]	
4d Modesty	2	3.55	0.85	.28	.10 [c]	.12 [c]	-.20 [b]	.16 [b]	.26	-.03 [c]	.00 [c]	-.04 [c]	-.14 [b]	.08 [c]	.56	.20 [a]	.16 [b]	.29

Note. N = 208. It. = number of scale/subscale items. All correlations are $p < .001$ except where noted. [a] = $p < .01$. [b] = $p < .05$. [c] = non-significant

Table 10

Hierarchical regression analyses of OCB on the factor- and facet level of honesty-humility and conscientiousness, and on integrity

	OCB				OCB		
Predictor (Factor level)	ΔR^2	β	*rw*	Predictor (Facet level)	ΔR^2	β	*rw*
Step 1	.16***			Step 1	.19***		
Conscientiousness		.33***	50.6%	Organization		.05	6.7%
				Diligence		.25***	26.0%
				Perfectionism		.05	5.9%
				Prudence		.06	5.7%
Step 2	.03**			Step 2	.10***		
Honesty-humility		.07	11.5%	Fairness		.25***	25.6%
				Greed avoidance		-.14*	4.0%
				Modesty		.06	2.1%
				Sincerity		-.05	1.0%
Step 3	.06***			Step 3	.04***		
Integrity		.27***	37.9%	Integrity		.22***	22.9%
Total R^2_{adj}	.24***			Total R^2_{adj}	.30***		

Note. N = 208. *p < .05; **p < .01; ***p < .001. *rw* = relative weight (contribution to R^2)

Only one of the four facets of conscientiousness significantly accounted for variance in OCB: the facet of diligence, β = 0.25, $t(196) = 3.51$, $p < .001$. With regard to honesty-humility, two of the four facets of honesty-humility significantly accounted for variance in job performance: the facet fairness, β = 0.25, $t(196) = 3.78$, $p < .001$ and the facet of greed avoidance, β = -0.14, $t(196) = -2.16$, $p < .05$.

When combining both predictive traits, the facets of conscientiousness and honesty-humility accounted for 77.0% of the total variance in OCB, whereas integrity accounted for 22.9% of the total variance in OCB. Comparable with this finding, the conscientiousness facet of diligence accounted for 26.0% and the honesty-humility facet of greed avoidance for 25.6% of the total variance in OCB.

Considering the principle of parsimony (Forster, 2000), the facet model explained a slightly larger amount of variance in OCB (R^2_{adj} = .24 vs. R^2_{adj} = .30). It is important to note that the facet model included more predictor variables at the expense of six degrees of freedom. Considering both, model fit and parsimony, the facet model is superior

(BIC = 949.30) to the factor model (BIC = 964.30), while a difference in BIC of 10 reveals a strong evidence (Raftery, 1995).

5.4 DISCUSSION

The key question in this chapter concerned the main constructs of the nomological network of integrity, particularly the relationship between integrity and OCB. This question referred to Hypotheses 1a and 1b. The data revealed that integrity and OCB are significantly related to each other on the factor level. On the facet level, three of four facets significantly correlated with OCB. Moreover, Hypothesis 2 examined the relationship between integrity and OCB while controlling for conscientiousness. The hypothesis that the relationship between integrity and OCB does not decrease by partialling out conscientiousness was confirmed by the results. Regarding the focus on the criterion, Hypotheses 3a and 3b investigated the incremental validity of integrity beyond the factors and facets of its important predictors of conscientiousness and honesty-humility when predicting OCB. Integrity revealed to be an important antecedent of OCB beyond both closely related personality traits on both the factor- and facet level of conscientiousness and honesty-humility. In the following section, all findings are discussed in detail and interpreted in light of the nomological network of integrity.

Relationship between integrity and OCB

The present findings demonstrate that integrity and OCB are two closely related concepts. In comparison with findings regarding the correlation between integrity and other working behaviors (CWB or job performance), the correlation between integrity and OCB (r = .37) is about equally high. Moreover, the correlation reflects the same high level as the correlations that were found for OCB with CWB, which is the key component of integrity (ρ = -.32 found by Dalal, 2005 up to ρ = -.39 found by Sackett et al., 2006).

According to Casillas et al. (2009), the correlation between integrity and OCB was found to be r = .70, which nearly twice the level in the current data (r = .37). There are three reasons for the assumption that Casillas et al. (2009) seem to have overestimated the correlation between integrity and OCB:

(a) As already mentioned, Casillas et al. (2009) failed to measure both integrity and OCB in a clear and separate manner. This means that the construct of integrity was measured with a very narrow focus on only two facets (general work attitudes and risk reduction) which does not represent the construct of integrity adequately. Moreover, the items of the integrity test were confused with OCB relevant content instead of referring to CWB. The construct of OCB was measured by generated supervisor performance ratings, so a well-established test to measure OCB was missing.

(b) Casillas et al.'s (2009) sample contained a large proportion of blue-collar workers, while the current results were mainly gained from white-collar workers. Blue-collar workers can be assumed on average to have more opportunities as well as need to show both OCB and integrity. For example, with regard to OCB, blue-collar workers operating on the same production line can easily help each other, thereby rendering it essential that they help one another because a pace must be kept. With regard to integrity, being lower in the hierarchy, blue-collar workers might have fewer opportunities to behave counterproductively than white-collar employees because of greater surveillance (e.g., by colleagues and the team leader) and tighter working conditions (e.g., pre-determined shifts).

(c) The magnitude of the correlation between integrity and OCB in the data at hand (r = .37) fits with other findings on the correlation of OCB with conscientiousness (r = .42 found by Miller, Griffin & Hart, 1999) and CWB (ρ = -.32 found in the meta-analysis by Dalal, 2005). Correlations of such a magnitude are only found when two subcomponents of one concept are related to each other (for example, .70 was found for interpersonal and organizational deviance in the meta-analysis by Dalal, 2005).

As conscientiousness is one of the most important personality antecedents to OCB (e.g., Borman et al., 2001; Podsakoff et al., 2000), the study at hand further demonstrated that integrity has been a long-neglected, albeit noteworthy correlate of OCB. Moreover, the study underlines the notion that OCB is also an important construct in the nomological network of integrity. Therefore, OCB has to be considered more strongly regarding the construct of integrity as well as the development of integrity tests.

With regard to the facet level, three of four OCB facets significantly correlated with integrity. The OCB facets of altruism, conscientiousness, and sportsmanship significantly correlated with integrity. Each of these three facets conceptionally relates to integrity and its behavior: (a) Altruism is an interactional behavior that includes helping other individuals

(colleagues, supervisors, customers) with work-related problems. (b) Conscientiousness as a facet refers to all behaviors that a good employee ought to perform, such as being parsimonious with working material and preparing working duties. (c) Sportsmanship is the ability to tolerate problems, stress, and strains in the working environment without becoming annoyed by them. This trait is in line with the finding that positive emotions lead to OCB, whereas negative emotions (i.e., anger or anxiety) lead to CWB (Spector & Fox, 2002). Therefore, positive emotions that have their source in the facet of sportsmanship prevent CWB.

The OCB facet of civic virtue did not significantly correlate with integrity, which is not surprising given the content of this subscale: Civic virtue refers to the organization on the macro perspective. Regarding the integrity of a single employee, it is not possible to act integer with reference to the entire organization except for special positions, such as experts or chief officers. The behavior of civic virtue relates to participation in actions to support and engage the complete organization, such as being informed about organizational business. Civic virtue might be an indirect indicator for integrity as opposed to a related behavioral concept.

Overall, the relationship between integrity and OCB appears stable because the correlation retains the same magnitude, even when conscientiousness—which is one of the most important correlates of integrity and OCB—is partialled out. This finding emphasizes the fact that other sub-concepts of OCB beyond conscientiousness support this strong relationship. Behavior in terms of the organization as well as behavior that supports other individuals or the rules of the organization is a main component of integrity.

In summary, the relationship between integrity and OCB is significant and stable. Nevertheless, the facets of OCB reveal a distinct pattern in their relationship with integrity. Hence, OCB and its facets of sportsmanship, conscientiousness, and altruism represent personality characteristics that have to be considered when conducting research on integrity, especially when sounding out the nomological network of integrity. Moreover and in line with the current study, “the prediction and understanding of behavior at work may be improved by focusing on more narrowly defined dimensions of integrity” (p. 151, Van Iddekinge et al., 2005).

Predicting OCB with integrity beyond conscientiousness and honesty-humility

The most important result regarding the prediction of OCB is the finding that integrity explains incremental validity beyond the factor and facets of both conscientiousness and honesty-humility (H3a and H3b). Moreover, integrity and honesty-humility explain as much variance in OCB as does conscientiousness.

It is remarkable that integrity (as well as honesty-humility) predicts OCB beyond conscientiousness because this trait has been consistently shown to be one of the best predictors of OCB (e.g., Borman et al., 2001; Ilies et al., 2009; Organ & Ryan, 1995; Podsakoff et al., 2000). Moreover, the finding that honesty-humility incrementally predicts OCB beyond conscientiousness is in line with prior studies (e.g., A. De Vries et al., 2011; M. K. Johnson et al., 2011). These studies have found that the addition of honesty-humility explains incremental variance in the prediction of performance criteria. Although honesty-humility has been found to be a well-examined predictor of OCB beyond the Big Five traits (M. K. Johnson et al., 2011), the study at hand found that integrity also explains additional variance in predicting OCB.

In contrast, a meta-analysis that investigated incremental validity by honesty-humility over diverse predictors (i.e., the Big Five, integrity tests, and general mental ability) found no gain in validity for predicting OCB (Y. Lee et al., 2019). This previous result and the result of the present study support integrity to outperform other predictors for predicting OCB.

The amount of explained variance produced by integrity as well as by honesty-humility beyond conscientiousness when predicting OCB is considerable when compared with incremental validity in previous studies on the predictors of OCB (e.g., Kluemper, DeGroot, & Choi, 2013). Moreover, previous studies dealing with the HEXACO model (including honesty-humility as well as conscientiousness) and integrity as predictors of CWB have found a comparable or greater gain in validity by integrity tests depending on which integrity test was used (Marcus et al., 2013). In general, the degree of explained variance is considerable because an increase in validity caused by a personality trait (i.e., integrity) beyond another personality trait (i.e., conscientiousness or honesty-humility) is unlikely, especially when both predictors have sustainable correlations to such a high degree. In contrast, the incremental validity of a personality trait beyond a non-personality factor (e.g., general mental ability) is greater. To give an example, a measure of individual-organizational value congruence was found to have only a slight incremental

validity for predicting four different types of work-related behavior (e.g., OCB) beyond the Big Five personality measure (ΔR^2 = .00 to .06) (Tsai, Chen, & Chen, 2012); in contrast, Schmidt and Hunter (1998) found a greater increase in incremental validity of integrity tests for predicting job performance beyond a test of general mental ability (ΔR^2 = .14).

Therefore, considering the incremental validity of various personality traits beyond conscientiousness or other personality traits in other studies on OCB (e.g., Bowling, 2010; Chiaburu & Lim, 2008; Kluemper et al., 2013), the incremental validity of honesty-humility and integrity beyond conscientiousness for OCB in the study at hand is considerable.

Regarding the facet model in detail, the conscientiousness facets of diligence and honesty-humility fairness had the highest beta weights (i.e., both β = .25). In terms of the conscientiousness facet of diligence, A. De Vries et al. (2011) found a similar pattern for the prediction of academic performance: The conscientiousness facet of diligence had the greatest significant gain (ΔR^2 = .18, p < .05). In comparison, the present findings also revealed that the conscientiousness facet of diligence produced the greatest significant gain (ΔR^2 = .25, p < .001). Regarding its content, the conscientiousness facet of diligence is related to OCB because people who work hard and in a disciplined manner also tend to support other people or the organization by doing extra work. Interestingly, the facet of diligence correlated greatest with integrity; nevertheless, integrity explains additional variance in the prediction of OCB beyond the facet of diligence. Moreover, the same result was found in Chapter 4 regarding the prediction of job performance. Thus, diligence seems to be an important facet regarding organizational behavior.

The finding of the honesty-humility facet of fairness is in line with prior studies that found the facet of fairness to be the most relevant personality facet concerning CWB (with a beta weight of 0.47) (R. E. De Vries & Van Gelder, 2013) and to incrementally predict CWB, which is contextual organizational behavior (Borman & Motowidlo, 1993) beyond justice perceptions (O'Neill et al., 2011). These findings are in line with the result of the current study, namely that the honesty-humility facet of fairness predicts OCB, which is also contextual performance. Fairness seems to be an important component of OCB because it is a precondition to helping other people or the organization in addition to doing extra work. Moreover, fairness revealed to be the honesty-humility facet that correlates the most with integrity (see Table 7 in Chapter 4). Nevertheless, integrity predicts OCB beyond the facet of

fairness, which is therefore an important correlate of organizational behavior referring to integrity.

In addition, previous findings have revealed that employees who perceive fairness in their organization (e.g., distributive and procedural justice) show more OCB (e.g., Cohen-Charash & Spector, 2001; Moorman, Blakely, & Niehoff, 1998). The fact that employees react to perceived fairness in organizations suggests that the honesty-humility facet of fairness is a personality variable that might enhance the connection of perceived fairness and OCB (e.g., Messer & White, 2006) or of distributive and procedural justice and OCB (e.g., Cohen-Charash & Spector, 2001). Specifically, people who score high on the facet of fairness might be more sensitive than people who score low on fairness in terms of spotting and appreciating fairness in their organization and in turn paying back experienced fairness in OCB.

The honesty-humility facet of greed avoidance had a significant negative gain in predicting OCB. This finding is in line with that of A. De Vries et al. (2011), who discovered that the facet of greed avoidance significantly predicts academic performance. As people with high levels of greed avoidance tend to strive for money, are uninterested in privilege and fame, and tend to distinguish themselves from others, this tendency is opposed to OCB, which would involve supporting others and being a helpful part of the group.

Theoretical implications for the nomological network of integrity

With regard to the findings in the current study, the concept of OCB is substantially related to integrity. The magnitude of this relationship is comparable with the correlation of integrity and CWB (Ones et al., 1993; Van Iddekinge et al., 2012a) as well as with the correlation of OCB and CWB (Dalal, 2005). While OCB and CWB are closely related by concept, the findings support the idea that integrity and OCB also have a strong conceptional relationship. Focusing on integrity tests that predict the occurrence of CWB, the question arises as to whether integrity tests refer to OCB.

Regarding the facet level of OCB, this question might be answered by the strong correlation between the OCB facets of altruism, conscientiousness, and sportsmanship (except for civic virtue) and integrity: Altruism is defined as voluntarily helping other people in the organization with a work problem (Organ, 1988, 1990a, 1990b). This description refers to integrity because this kind of behavior seems to be contrary regarding CWB, which refers

to other individuals in the organization. Moreover, altruism might reduce CWB in other people who are helped; therefore, altruism is a component of the conglomerate of integrity.

Conscientiousness refers to behaviors that are intended to benefit the organization. The central point is that this facet describes "what a good employee ought to do" (C. A. Smith, Organ, & Near, 1983, p. 657). Therefore, Organ and Ryan (1995) referred to the generalized compliance dimension of OCB, and Podsakoff et al. (2000) established generalized compliance as one of seven main OCB categories found in the OCB literature. With regard to integrity, compliance seems to be central because employees who respect the internal and external rules and norms of an organization act with integrity. This component is the key point that describes integrity and what integrity tests should aim to measure.

Sportsmanship is defined as having "a willingness to tolerate the inevitable inconveniences and impositions of work without complaining" (Organ, 1990b, p. 96). With regard to integrity, sportsmanship seems to be more than a simple attitude. In light of integrity, sportsmanship seems to be a stable basic characteristic of people who are dealing with adverse conditions. This idea is important because of rationalization mechanisms as people who are confronted with hard circumstances tend to justify CWB (Fine & Gottlieb-Litvin, 2013; Greenberg, 1990). Due to the significant relationships between integrity and the three OCB facets, it seems necessary to widen the definition of integrity by including OCB and some of its facets.

Regarding the facet level of the predictors, both the conscientiousness facet of diligence and the honesty-humility facet of fairness display the largest in explained variance when predicting OCB. Even though both facets are related to integrity, integrity incrementally predicts OCB beyond these two facets. Therefore, this finding strengthens the idea that both facets have to be kept in mind regarding the nomological network of integrity.

Future Research

Future research should aim to strengthen the current findings on integrity and OCB by using different concepts of OCB and integrity tests, thereby widening the general validity of both constructs:

First, other concepts of OCB than those based on Organ's model (1988) could be used. About 30 different forms of OCB have been identified in the literature (Podsakoff,

MacKenzie, Paine, & Bachrach, 2000). The taxonomy of OCB-I (i.e., OCB against individuals) and OCB-O (i.e., OCB against an organization) (Williams & Anderson, 1991) is synonymous with the classification of CWB that targets individuals and the organization (Robinson & Bennett, 1995). This observation could provide a narrower focus on the relationships between all variables.

Second, comparing overt and personality-based integrity tests helps to more comprehensively elucidate the link between integrity and OCB. Moreover, the incremental validity of predicting OCB via integrity may change due to the use of a personality-based tests instead of an overt integrity test: Regarding the fact that overt integrity tests better predict CWB than do personality-based tests, personality-based measures are expected to provide a lower prediction of the criterion of OCB. Future findings from different integrity tests and the relationship between different concepts of OCB could shed more light on the nomological network of integrity and its construct.

Third, other personality traits might be controlled for when incrementally predicting OCB via integrity. The personality characteristic of agreeableness and its facets are importantly related to both OCB (Chiaburu et al., 2011) and integrity (Sackett & Wanek, 1996). Moreover, integrity is defined as a compound trait of agreeableness, conscientiousness, and emotional stability (Berry, Sackett, & Wiemann, 2007). Therefore, emotional stability might be another candidate to control for when examining the prediction of OCB by integrity (Chiaburu et al., 2011; Singh & Singh, 2009). In contrast to the Big Five characteristics, another promising trait to control for might be self-control. Self-control was found to be an important correlate of integrity (Bazzy et al., 2017). Moreover, self-control also incrementally predicted CWB beyond age and situational variables (Marcus & Schuler, 2004). In this case, it would be interesting to investigate whether integrity shows incremental validity for predicting contextual performance, such as CWB and OCB.

Strengths and limitations

As already mentioned in the discussion of Chapter 4, the data and measure of the variables comprise some strengths and restrictions.

The demographic and occupational characteristics of our sample go hand in hand with both challenges and opportunities for the investigation: On the one hand, the data sample was very specific (68% of the participants were women who were very young with a

mean of 26 years, relatively well-educated, and predominantly working part-time). Nevertheless, other studies have also included specific samples and revealed results that fit with the present findings (Bourdage et al., 2012; Feather & Rauter, 2004). On the other hand, our sample referred to a huge variation of occupations, which supports the generalizability of the present findings.

The use of self-reports by the same person to assess all variables is critical because methodical bias might occur (Podsakoff et al., 2003). Nevertheless, other studies have revealed that ratings of others do not highly correlate with self-ratings (e.g., Allen, Barnard, Rush, & Russell, 2000; Khalid & Ali, 2005) and that self-ratings are comparable (e.g., K. Lee & Ashton, 2013) and can even be more accurate (Ashton & Lee, 2010) than observer reports.

The ongoing problem of integrity measures' tendency to be vulnerable to social desirability (Alliger & Dwight, 2000) is reduced by using an anonymous web-based setting (Crutzen & Göritz, 2011). Moreover, because all used variables are prone to socially desirable answers, the effect appears with all variables. All values may thereby increase, but the effects regarding the present findings (correlations and regressions) should remain the same.

Conclusion

The present study investigated the hitherto neglected relationship between integrity and OCB. Integrity is significantly related to OCB and its facets of altruism, conscientiousness, and sportsmanship. Moreover, integrity explains additional variance beyond that of the best-established, investigated personality trait of OCB—conscientiousness—and even beyond honesty-humility at the factor- and facet level. In particular, the conscientiousness facet of diligence as well as the honesty-humility facet of fairness explained additional variance in OCB beyond integrity, with which the facets are related. Therefore, OCB as well as the three facets might henceforth be considered in the nomological network of integrity. Moreover, OCB as well as its three facets are related in their content and are therefore necessary components of the construct of integrity in addition to the personality facets of diligence and fairness. Finally, the findings should encourage future research on the concept of OCB with reference to the conglomerate construct of integrity.

CHAPTER 6 EVALUATING TWO PROMISING METHODS TO REDUCE FAKING ON INTEGRITY TESTS

One of the main unresolved concerns of integrity tests is response distortion (faking) among applicants (e.g., Alliger & Dwight, 2000; Berry, Sackett, & Wiemann, 2007; Morgeson et al., 2007; Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett & Christiansen, 2007). Although many approaches to preventing response distortion on integrity tests have been developed (e.g., Jackson et al., 2000; McDaniel et al., 2007), faking has not yet managed to be effectively prevented (Karren & Zacharias, 2007). Therefore, two rating methods for preventing faking on integrity tests were tested: the double-rating method developed by Hui (2001) and the indirect questioning method (e.g., Fisher ,1993). Moreover, the underlying mechanism of both methods was further investigated, and a simulated personnel selection setting was exploratively examined.

6.1 THEORY AND HYPOTHESES

Faking has often been investigated in the context of diagnostic testing, especially in employee selection (e.g., Ellingson, Sackett & Connelly, 2007; Griffith & McDaniel, 2006; Schmitt & Oswald, 2006). Applicants tend to put a positive spin on their answers or behavior in order to strengthen the appearance of preferable personality characteristics, such as conscientiousness and emotional stability (e.g., Gerber-Braun, 2010; Griffith et al., 2007).

As a result, faking threatens the validity of different measures , as has been revealed by numerous studies (e.g., Caldwell-Andrews, Baer, & Berry, 2000; Dilchert, Ones, Viswesvaran, & Deller, 2006; Griffith & Peterson, 2008; Tett & Christiansen, 2007; Zickar & Drasgow, 1996). Moreover, one meta-analysis found that the scores of an integrity measure significantly improve when applicants distort their answers (Alliger & Dwight, 2000). In another meta-analysis, different traits from one personality scale were found to be influenced by applicant faking, with an effect size ranging from d = .11 to d = .45 (Birkeland et al., 2006). Applicants who fake thus receive higher values for preferable personality traits and therefore have a greater likelihood of being hired for a job. Faking can thereby significantly affect the ranking of qualified applicants (Birkeland et al., 2006; Griffith et al., 2007; Peterson et al., 2009).

Integrity tests are especially vulnerable to faking because they explicitly test for desirable attitudes, personality traits, and admitted minor wrongdoings. Research has indicated that applicants can fake (e.g., Viswesvaran & Ones, 1999) and do fake (e.g., Dwight & Alliger, 1997; Griffith et al., 2007) on integrity tests in personnel selection settings (e.g., Birkeland et al., 2006; Marcus, 2006) and even within simulated selection settings (e.g., Gerber-Braun, 2010; Jackson et al., 2000). In addition, McFarland and Ryan (2000) found that an integrity test is one of the measures that is most often faked. Indeed, the results of an integrity test are significantly better for the faking condition (i.e., participants are instructed to fake the test) than for the honest condition (i.e., participants are instructed to answer honestly). One meta-analysis revealed that due to faking, integrity test values shift by about one-half and up to one standard deviation and that in contrast to personality-based integrity tests, overt integrity tests are particularly susceptible to faking (Alliger & Dwight, 2000).

Different methods have been developed to address the problem of faking, but doubts about the effectiveness of these methods persist for both empirical (e.g., Peterson, Griffith, O'Connell, & Isaacson, 2008) and ethical reasons (e.g., for the bogus pipeline method). For example, researchers agree that neither validation subscales nor social-desirability scales can effectively account for the variance caused by faking (e.g., Griffith & Peterson, 2008; Peterson et al., 2011; Sackett, 2011).

As a result of these findings, Burns and Christiansen (2011) stated that no sufficiently successful method for preventing applicant faking yet exists. The problem of faking persists for personnel selection and its methods (e.g., Morgeson, 2004). In a meta-analysis, Birkeland et al. (2006) found that there is a "need to continue investigating ways to reduce applicant score distortion" (p. 328). Moreover, in their review Karren and Zacharias (2007) recommend that future research focus on methods to reduce faking on integrity tests.

Identifying and further developing useful approaches to combating faking is thereby essential. As most of these methods aim at combating faking after it has already occurred, it is important to concentrate on methods that prevent socially desirable responding before it occurs.

In social psychology, Hui (2001) introduced a promising method that is easy to apply and can prevent faking before it occurs. In a first step, Hui (2001) instructed his participants to fill out a social desirability questionnaire by providing responses that they would expect others to give ("Which answers do you think people would choose in order to appease

others?", labelled *other-rating)*. In a second step, Hui (2001) asked his participants to fill out the same questionnaire again and to decide how well the statements fit with their own opinion ("How accurately do the statements describe your own behavior and attitudes?", labelled *self-rating*). Hui (2001) labelled this two-step rating procedure the 'double-rating method'.

In his study, Hui (2001) had two conditions for ratings: The control group filled out the questionnaire with their own opinions, and the experimental group filled out the questionnaire two times: first with their beliefs about the opinions of others and second with their own opinions. Hui (2001) found that the self-rating of social desirability in the experimental group (self-rating in the double-rating method) was lower than the self-rating of social desirability in the control group (self-rating only).

Hui (2001) argued that the participants had answered more truthfully on issues of undesirable behavior or attitudes if they had felt comfortable and safe and if their responses had supported a positive self-image. If participants had believed that they were in the majority, they had likely felt better about themselves and judge themselves less harshly. Therefore, Hui (2001) concluded that the participants' self-image had remained positive if they had believed that undesirable behaviors or attitudes were common. Participants who had expressed that their own behaviors or attitudes were common in the other-rating had legitimized their own behaviors or attitudes. Moreover, Hui (2001) argued that the participants' self-image had grown more positive than when they had assumed that they were in the minority.

Despite this rudimentary reasoning, Hui (2001) did not provide any theory or theoretical effect as a basis for his method. He thus set out to test his assumptions and to investigate the effectiveness of his method via two experiments. In both cases, he found that participants who had provided both ratings (i.e., the other-rating and the self-rating) scored significantly lower on the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1964) than did the control group, whose members had only responded with their own opinions (Hui, 2001). In addition, participants' answers in both the self- and the other-rating showed a significant positive correlation. These findings indicate that the double-rating method can decrease socially desirable responding by bringing the occurrence of others' undesirable behaviors or attitudes to the participants' attention.

Thomas et al. (2007) replicated and extended Hui's (2001) double-rating method. In addition to the Marlowe-Crowne Social Desirability Scale (MCSDS, Crowne & Marlowe, 1964) used by Hui (2001), Thomas et al. (2007) included a second scale of social desirability, the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984, 1988). This scale was added for two reasons: (a) its two subscales (self-deception and impression management) and (b) its more-differentiated response format (7-point Likert scale) in comparison with the yes/no response format of the MCSDS.

The results of Thomas et al. (2007) supported Hui's results obtained with the MCSDS: Thomas et al. (2007) found that participants in the single-rating condition had higher social desirability scores than did participants in the self-rating of the double-rating condition for both the MCSDS and the BIDR. Moreover, the correlation between the other-rating of social desirability and the self-rating of social desirability was significant and positive for both social desirability scales. No significant difference was found between the self-rating and the other-rating within the double-rating method for the BIDR or between either of its subscales. This result supports Hui's (2001) explanation of the mechanism of the double-rating method. Finally, Thomas et al. (2007) concluded that while the mechanism of the double-rating method is not fully clear, the method itself does appear to reduce socially desirable responding.

Given its potential to reduce faking, it is necessary to further test ability of the double-rating method to reduce socially desirable responding within constructs and for measures other than social desirability. As integrity tests are especially prone to socially desirable responding (Alliger & Dwight, 2000; McFarland & Ryan, 2006), the double-rating method should be tested in terms of its ability to reduce faking on integrity tests. Moreover, the recommendations given by Hui (2001) and Thomas et al. (2007) concerning the mechanism of the double-rating method should be considered. Therefore, the current study broadens the field of research on the double-rating method and applies it to integrity tests. Furthermore, it contributes to a better understanding of the mechanism that underlies the double-rating method.

Similar to the double-rating method, which includes a rating from the perspective of others in its first step, indirect questioning is another method that consists of an other-rating and that has the potential to reduce faking (Fisher, 1993; Dalal & Hakel, 2016). While the double-rating method asks participants to answer based on their idea of what others might

think in addition to answering for themselves, the method of indirect questioning only asks participants to answer from the perspective of some average other person. Fisher (1993) assumed that participants project their beliefs and evaluations when answering indirect questions and found that indirect questioning can reduce the social-desirability bias: Using indirect as opposed to direct questioning (i.e., a rating of one's own attitudes or behaviors), he found that participants provided higher scores for a construct that was socially desirable than for a non-sensitive construct. Using a non-socially desirable construct, Fisher (1993) found no difference between answers in the indirect and direct questioning methods.

In terms of the vulnerability of integrity tests to socially desirable responding and the ineffectiveness of existing methods at reducing faking, the present study is important for three main reasons: (a) With regard to the double-rating method, the study offers a widely unknown yet promising approach from social psychology for reducing faking on integrity tests. (b) Both approaches (i.e., the double-rating method and indirect questioning) can prevent faking before it occurs and are therefore helpful when it comes to reducing faking. (c) The study contributes to a basic understanding of the self- and other-ratings on integrity tests. Several variables were added that contribute to a basic understanding of both the double-rating method and indirect questioning. Hence, regarding the operationalization and the aim of the current study, namely to make a first attempt at testing novel approaches that could prevent faking on integrity tests, several hypotheses were proposed.

Both Hui (2001) and Thomas et al. (2007) found that the double-rating method can reduce socially desirable answering on measures of social desirability because the self-rating in the double-rating method in their studies was lower for social desirability than was the self-rating only of social desirability. With regard to the method of indirect questioning, three studies on the construct of CWB revealed that more CWB was reported when using indirect questioning than when using other methods for reducing socially desirable answering (e.g., anonymity) (Dalal & Hakel, 2016).

As a result of these findings and the fact that no method for efficiently preventing faking on integrity tests yet exists, the ability of both the double-rating method and the indirect questioning method to reduce faking on integrity tests needed to be tested. Moreover, rating formats other than the standard rating of tests (i.e., a 5-point self-rating scale)—such as a forced-choice answer format—have been shown to be capable of reducing faking (Jackson et al., 2000). Therefore, in the present study, both the double-rating method

and indirect questioning were included in addition to a control group with a self-rating only condition (for an overview of experimental conditions, see Appendix, Table B).

To demonstrate the effect that both measures could have on reducing socially desirable responding, two approaches were used for data analysis: Previous studies mostly used mean differences in the ratings of different faking-reducing methods to demonstrate a method's effectivity at reducing response distortions (e.g., Dalal & Hakel, 2016; Tourangeau & Yan, 2007). In the present study, the approach of mean differences was therefore used for both the double-rating method and indirect questioning. Moreover, some studies used a correlational approach to demonstrate the effectivity of methods at reducing faking (e.g., Dalal & Hakel, 2016). In the present study, the correlational approach was additionally used for the double-rating method. I thus propose the following hypotheses:

H1: The self-rating of the double-rating method is lower for integrity than is self-rating only (control group).

H2: The difference between the self- and the other-rating within the double-rating method positively correlates with impression management.

H3: The other-rating of the indirect questioning method is lower for integrity than is self-rating only (control group).

For the double-rating method, Hui (2001) found that self-rating was higher than other-rating for socially desirable responding. He interpreted this difference between the self- and the other-ratings as evidence that participants answer more truthfully if they provide a high rating for an undesirable behavior or attitude of others. Additionally, he concluded that participants who themselves have a tendency to show an undesirable behavior or attitude feel like they are part of a majority when giving others a high rating on the tendency to show this undesirable behavior or attitude. Moreover, Hui (2001) stated that this effect does not hold when participants believe that they belong to the minority, but he failed to test this assumption empirically. There could be other effects that contradict this explanation, and Hui's (2001) assumption therefore requires testing. To clarify Hui's (2001) position—namely that participants feel a sense of belonging to a majority, which is the key

component of the effect of reducing socially desirable responding in the double-rating method—two different reference groups were created for the other-rating: the reference group of the majority and the reference group of the minority. In the majority condition, participants referred to people in common with the integrity-related other-rating—that is, participants were given an integrity-related statement and were asked, "To what extent does this statement apply to others?" With the condition of the majority, participants were led to believe that they were part of the majority regarding their integrity-related attitude or behavior. In contrast, in the minority condition, participants referred to one person only (who was described in a particular situation) with their integrity-related other-rating–that is, participants were asked, for example, "To what extent does this statement apply to Mrs. Smith?" The condition of the minority let participants to refer only to one person and therefore to believe that they were part of a minority in terms of their integrity-related attitude or behavior. As a result, if both reference groups differ in the integrity scores for the other-rating, Hui's (2001) explanation is supported. Therefore, I propose the following hypothesis:

H4: The difference in self-ratings between the double-rating method and the self-rating only (control group) is greater in the majority condition than in the minority condition.

Both reference groups (the majority and the minority) were additionally tested using the indirect questioning method. Using the indirect questioning technique, participants project their beliefs and evaluations while responding to their perception of what others might do or think. In line with Hui's (2001) explanation of the double-rating method, in the indirect questioning method, participants might also indicate that they have lower integrity because they believe that they are part of the majority. In contrast, it might also be possible for participants to more strongly refer only to a single person (minority) with their integrity rating because they identify with that person (e.g., a person who is part of an in-group). As a result, the reference group of the minority might provide a better projection of their own attitudes and behavior. To investigate this issue, the following research question was included:

RQ: Do the other-ratings of the indirect questioning method yield a different integrity score for the reference group of the majority in contrast to the reference group of an individual (minority)?

The fact that integrity tests are especially relevant for personnel selection and that this setting provides both the motivation and the opportunity to fake also needs to be taken into account. Studies have emphasized that scores on integrity tests increase considerably in real personnel selection procedures (Marcus, 2006; Van Iddekinge, Raymark, Eidson, & Putka, 2003). Even simulated selection settings are prone to faking (Gerber-Braun, 2010; Jackson et al., 2000; Marcus, 2006). As a result, faking has substantial effects on rank order and hiring decisions (Griffith et al., 2007). Keeping in mind the fact that the double-rating and the indirect questioning method both aim to reduce faking, the question arises as to how these methods interact in a personnel selection setting. To explore this question in a first approach, the condition of a simulated personnel selection setting was integrated in order to conduct an explorative analysis.

6.2 METHODS

Participants

Data were collected from the WiSo-Panel, a nonprofit, university-based online panel from Germany with demographically heterogeneous participants. A total of 5,852 panelists were invited and randomly assigned to an experimental or control group. The response rate was 27%, and the retention rate was 25%, thereby enabling a dataset of $N = 1{,}450$ participants to be used in the study. Participants had an average age of $M = 44$ years ($SD = 10.70$), ranging from 20 to 77 years. Fifty-seven percent of the sample consisted of women ($n = 833$). Most of the participants had a university degree (33.1%), followed by a middle school degree (29.9%) or a university-entrance qualification (24.0%); only few participants had a secondary-school degree (9.8%), a doctoral degree (3.0%) or no school degree (0.2%). The work experience of the sample was on average $M = 26$ years ($SD = 14.79$). A total of 95.2% of participants from the sample held jobs, and 16.8% of participants were in the process of applying for a job.

Measures

Impression management. The 10-item impression management scale, which was used in the current study, was taken from a German version of the Balanced Inventory of Desirable Responding (BIDR), which was developed by Musch, Brockhaus, and Bröder (2002). This German version of the BIDR is in line with the original BIDR (Paulhus, 1988), which uses two factors of social desirability, namely, self-deceptive enhancement and impression management (Paulhus, 1984, 1994). While self-deceptive enhancement involves presenting oneself in a better light by unconsciously deceiving oneself, impression management involves presenting oneself in a better light by consciously deceiving others. In contrast to self-deceptive enhancement, the impression management subscale is sensitive to situations relating to the motive of good self-presentation when dealing with others (e.g., "I never swear"). Therefore, impression management is appropriate to control for deception within the personnel selection procedure (Musch et al., 2002). Convergent and discriminant validity have been demonstrated in both dimensions of the BIDR in comparison with other measures of social desirability (e.g., Skala zur Messung Sozialer Erwünschtheit from Mummendey & Eifler, 1993 or Soziale-Erwünschtheitsskala-17, SES-17, from Stöber, 1999). The two-component structure of the BIDR was demonstrated via results of non-correlation for both subscales, via a two-factor solution using an exploratory factor analysis, and via a different correlational pattern of both subscales with personality scales (Paulhus, 1984; Paulhus, 1994). Both the original BIDR and its German version use a 7-point answer format ranging from 1 (*strongly agree*) to 7 (*strongly disagree*). For the impression management scale, the test-retest correlation was r_{tt} = .65 (Paulhus, 1991). The internal consistency of the impression management scale ranged from α = .75 to .81 (Paulhus, 1991). In the present study, the internal consistency of impression management was α = .67.

Integrity items. Six integrity items were developed and pre-tested (see Appendix, Table C). All items were based on a dilemma situation in which an actor had to decide whether to show integrity or CWB. Counterproductive actions that are common in business, such as time misuse or private use of communication media, were chosen (Gruys & Sackett, 2003). Because items were based on widespread CWB, the problem of having CWB items that did not pertain to some participants was eliminated (Bowling & Gruys, 2009). Every item was designed to include a described situation for two reasons: First, situational judgement tests are able to reduce socially desirable answering (Kasten, Freund, & Staufenbiel, 2020;

McDaniel et al., 2007) and can therefore be effectively combined with new methods of faking prevention (e.g., Oostrom, Köbis, Ronay, & Cremers, 2017). Second, the integrity-related situations that were created for this study always involve small wrongdoings that can take place at a company. By using these situations (e.g., a person writes a private email at work), higher frequencies of counterproductive behavioral ratings could be expected to be reported by the participants in contrast to situational items involving a serious crime at a company (e.g., stealing company secrets). Moreover, a greater variation in the occurrence of CWB could be given because participants are more inclined to admit to small wrongdoings opposed to serious delinquencies. After reading the situational description, participants agreed or disagreed with a given statement on a 5-point Likert scale ranging from 1 *(applicable)* to 5 *(not applicable)* (e.g., "I would not take care of private issues at work."). In the current study, the internal consistency for the integrity items was $\alpha = .70$.

Manipulation checks

Control variables. Manipulation checks for the experimental conditions were performed. Every rating was controlled in terms of the person for whom the participants had answered the items (e.g., oneself, the person described in the situation, a colleague, a friend). Moreover, manipulation control was performed if the participants identified with the person described in each of the six integrity-related situations. Both control variables were added in order to gain a better understanding of the mechanism of the double-rating method and its effectiveness at reducing faking.

With regard to the explorative analysis of the simulated personnel selection setting, the quality of the simulated selection procedure was controlled with one question ("How easy would it be for you to put yourself in the position of the specified job application situation?"), which included a 7-point answer scale ranging from 1 *(very difficult)* to 7 *(very easy)*.

Procedure

A web-based inquiry with a 2 (method: double-rating and indirect questioning) x 2 (reference group: majority and minority) + 1 (control group of a self-rating only) x 2 (situation: anonymous survey and simulated personnel selection) between-subjects experimental design was conducted. Half of the participants were assigned to an anonymous

setting, and the other half were assigned to a simulated personnel selection setting for exploratory reasons. For the simulated personnel selection situation, the manipulation was introduced to the participants by two text passages: First, participants were prompted to imaging that they had applied for a job and now take an online-assessment for the applied job. Second, participants were welcomed as applicants for the online-assessment selection process.

Participants from the survey platform WiSo-Panel were invited via email. Demographic data (age, gender, education, and profession) were collected in a master file that contained a collection of stored data on every member of the survey panel. The survey began with a short greeting and an explanation of the process. First, participants answered some questions about their current professional status (work experience, current working status, and current job-application status). Afterward, they were randomly assigned to one of the experimental conditions for answering the six integrity-related items: In the experimental conditions, the method (the double-rating method vs. indirect questioning plus the control group of a self-rating only), and the reference group (majority vs. minority) were varied.

For the double-rating method, the order of the self- and the other-rating was counterbalanced between participants. Participants determined their self-rating and the other-rating for a single item before moving on to the next item. In contrast to this procedure, participants in Hui's (2001) experimental design first filled out the complete questionnaire for the other-rating and then completed the questionnaire for the self-rating. After presenting all integrity-related items, manipulation control was done. Depending on the condition, participants were asked (a) whom they had in mind while answering, (b) whether they identified with the person in each of the six situations described, and/or (c) whether they identified with the simulated personnel selection situation. At the end of the survey, participants filled out the 10-item impression management scale from a short form of the BIRD. In the survey, items of all measures were presented to participants in randomized order.

6.3 RESULTS

The following section begins with a preliminary data analysis designed to prepare data for further analysis in terms of the key features of the experimental groups. The section continues with tests of all hypotheses and the research question. The key issues deal with the potential to reduce faking via both the double-rating method and indirect questioning and with the underlying mechanisms of both methods. The section continues with a data analysis of the control variables. To conclude, an exploratory analysis is presented that investigates how the means of both the double-rating method and indirect questioning interact in a simulated personnel selection setting.

Preliminary data analysis

The first analysis involves the reference groups of the majority and minority for both the double-rating and the indirect questioning method. These two reference groups were created in order to investigate the underlying mechanism of both rating methods and are in no way related to the primary question in this study, which involves reducing faking on integrity tests. The difference between both reference groups was therefore tested preliminarily because if no significant difference can be found, it would be possible to aggregate data for the main questions of this study.

Two ANOVAs were conducted to test the difference between the two reference groups for both rating methods: one for the double-rating method and one for the indirect questioning technique. Homogeneity of variances was asserted using Levene's Test, which revealed that equal variances could be assumed (p = .186 for the double-rating method and p = .716 for indirect questioning). Results for double-rating method revealed that the means of both reference groups did not significantly differ, F (1, 409) = 0.26, p = .608, partial η^2 = .001. Results for the indirect questioning method revealed that the means of both reference groups also did not significantly differ, F (1, 215) = 0.51, p = .477, partial η^2 = .002. As a result, the self-ratings in the double-rating method and the other-ratings in the indirect questioning method can be summarized across both reference groups to test the hypotheses regarding the potential of rating methods to reduce faking.

Tests of hypotheses and research question

The integrity score for each condition is shown in Table 11. The first hypothesis to be tested posited that the self-rating of the double-rating method should be lower for integrity than should be self-rating only. An ANOVA was conducted. Homogeneity of variances was asserted using Levene's Test, which revealed that equal variances could be assumed (p = .348), meaning that the variances of all groups did not differ. The result revealed that there is a difference in integrity test scores between the self-rating of the double-rating method and the self-rating only, $F(1, 515) = 3.87$, $p < .05$, partial $\eta^2 = .007$. The effect size indicated that the effect is low (Cohen, 1988). These result supported H1.

In addition to comparing means, H2 postulated a significant correlation between the difference of the self- and the other-rating in the double-rating method for impression management. Results revealed that the correlation was significant and positive, $r = .41$, $p < .001$. Therefore, H2 was supported. This finding implies that the greater the difference was between both ratings (i.e., the self-rating minus the other-rating) within the double-rating method, the more that impression management was shown. With reference to Cohen (1988), this correlation represents a moderate effect.

To test H3, which posited that indirect questioning should be lower than the self-rating only, an ANOVA was conducted. Homogeneity of variances was asserted using Levene's Test, which showed that equal variances could be assumed (p = .258). Results revealed that there was no difference in integrity test scores between indirect questioning and the self-rating only, $F(1, 321) = 2.55$, $p = .111$, partial $\eta^2 = .008$. The results thus did not support H3.

Table 11

Means and standards of the self-ratings and the other-ratings

Condition	Anonymous survey					Simulated selection procedure				
	n	Self-rating		Other-rating		*n*	Self-rating		Other-rating	
		M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*
Self-rating only (control group)	106	18.95	4.81			101	19.66	4.59		
Double-rating majority	205	18.09	4.74	16.16	4.04	204	19.06	3.95	16.56	3.72
Double-rating minority	206	17.86	4.26	16.44	3.30	201	19.09	4.23	16.87	3.47
Indirect questioning majority	107			17.87	4.51	105			19.30	4.49
Indirect questioning minority	110			18.30	4.40	105			18.97	4.52

Note. N = 1,450 (n of groups and aggregated groups are added). Means and standard deviations for both reference groups of both rating methods are presented to provide a more-detailed overview. Scores for the explorative analysis of the simulated selection setting are also included to provide a complete and simultaneous overview of all means and standard deviations.

The next hypothesis involved Hui's (2001) explanation of the underlying mechanism of the double-rating method. H4 stated that the self-rating in the double-rating method should be lower for integrity when the other-rating of the double-rating method includes the majority instead of the minority as the reference group. To test this hypothesis, effect sizes of Cohen's *d* were calculated for the difference between the self-rating only (control group) and the self-rating of the double-rating method with the reference group of majority, d = -.18, and for the difference between the self-rating only (control group) and the self-rating of the double-rating method with the reference group of the minority, d = -.25. In the next step, the confidence intervals of both effect sizes were calculated: [-.42 to .05] for the difference of the reference group of the majority and [-.48 to -.01] for the difference of the reference group of the minority, both at confidence intervals of 95%. Results revealed that both confidence intervals overlap, which indicates that there is no significant difference for the double-rating method between (a) the difference between the control group and the

reference group of majority and (b) the difference between the control group and the reference group of minority. Therefore, H4 was not supported.

The research question investigated whether the other-rating of the indirect questioning technique yields a different integrity score for either reference group (majority and minority) (RQ). This question was also tested by comparing the difference between the other-rating of the indirect questioning technique of both reference groups and the control group of the self-rating only. Effect sizes of Cohen's *d* were *d* = -.23 for the reference group of the majority and *d* = -.14 for the reference group of the minority. The confidence intervals of both effect sizes were [-.50 to .04] for the difference concerning the reference group of the majority and [-.41 to -.13] for the difference concerning the reference group of the minority, both of which had confidence intervals of 95%. Results revealed that both confidence intervals overlap, which indicates that there is no significant difference between either of the two differences, that is, those of the control group and the reference groups of the majority and the minority for the indirect questioning technique. It can thus be concluded that neither reference group differs with respect to the indirect questioning technique.

Analysis of control variables

The analysis of the control variables yielded further insights into the mechanism behind the methods. The first control variable investigated the reference group to which the participants referred to when answering the six integrity-related items. As expected, almost all participants referred to themselves when completing a self-rating only or a self-rating within the double-rating method (about 90%). In addition, 21% of participants claimed to have referred to themselves when answering the other-rating within the double-rating method.

With regard to indirect questioning, a different pattern was found: Only about two-thirds of participants claimed to have referred to themselves when answering the other-rating of the indirect questioning technique (70%). The remainder of the participants claimed to have referred to other individuals (13%) or to both themselves and others (18%).

The next control variable concerned the issue of identification and investigated whether participants identified with the person in the situations described by the items. Results revealed that participants of all groups gave an average rating of 1.76 (*SD* = 0.64) in a

3-point answer scale with the three options of *yes* (indicated by a score of 1), *partially* (indicated by a score of 2), and *no* (indicated by a score of 3). Moreover, 89% of participants from all groups claimed to have identified or partially identified with the person described in the situation.

Explorative data analysis of simulated personnel selection setting

The simulated personnel selection procedure controlled for whether participants could imagine the simulated setting. Results reveal that participants gave a rating of M = 4,87 (SD = 1,42) on a 7-point rating scale ranging from 1 (*very bad)* to 7 (*very good)*. All groups of the simulated setting were designed in the same way as the experimental groups of the anonymous survey setting (i.e., both settings comprised (a) the control group of the self-rating only, (b) groups for the double-rating method (including the reference groups of the majority and of the minority), and (c) groups for the indirect questioning method (including the reference groups of the majority and of the minority).

Regarding the key issue of this study—namely to find a method of reducing faking on integrity tests—the explorative analysis of the simulated setting focused on analyzing all means of both methods (i.e., the double-rating method and the indirect questioning method) and comparing them with the condition of the anonymous survey setting. Moreover, within the simulated setting, means of both the double-rating method and the indirect questioning method were compared with the control group using the self-rating only method. Therefore and in line with the preliminary analysis of the anonymous survey setting, both reference groups of the double-rating method were tested to determine whether they differed significantly. Levene's Test was used to investigate the homogeneity of variances. Results revealed that equal variances could be assumed (p = .225), and that there was no significant difference between either reference group concerning the double-rating method for the simulated personnel selection setting, F (1, 403) = 0.00; p = .949, partial η^2 = .000). As a consequence, scores for both groups of the double-rating method were aggregated for further analyses in terms of their reduction of faking within the simulated setting.

Further analyses were conducted to investigate whether there was a difference between the double-rating method and the control group of self-rating only for the simulated personnel selection setting. Again, Levene's Test was used to test the

homogeneity of variances. Results revealed that equal variances could be assumed (p = .129) and that there was no significant difference between the double-rating method and the self-rating only condition for simulated personnel selection setting, F (1, 504) = 1.58; p = .209, partial η^2 = .003).

Regarding the method of indirect questioning and the preliminary analysis of the anonymous survey setting, both reference groups of the indirect questioning method were also tested to determine whether they differed significantly. Homogeneity of variances was determined using Levene's Test, which revealed that equal variances could be assumed (p = .989). Results revealed that there was no significant difference between either reference group for the indirect questioning method relating to the situation of the simulated personnel selection setting, F (1, 208) = 0.27; p = .603, partial η^2 = .001). As a result, both groups of indirect questioning were summarized for further analyses in terms of the reduction of faking within the simulated setting.

Further analyses of the indirect questioning technique explored possible differences between the method of indirect questioning and the control group using self-rating only for the simulated personnel selection setting. Homogeneity of variances was asserted using Levene's Test, which revealed that equal variances could be assumed (p = .896). Results revealed that there is no significant difference between the indirect questioning method and self-rating only for the situation of the simulated personnel selection setting, F (1, 309) = 0.935; p = .334, partial η^2 = .003).

The next analysis investigated whether the double-rating method scores lower on integrity when participants' integrity-related answers refer to the reference group of the majority as opposed to the minority. To test this question, effect sizes of Cohen's d were calculated for the difference between self-rating only (control group) and the double-rating method with the reference group of the majority, d = -.14, and for the difference between the self-rating only (control group) and the double-rating method with the reference group of the minority, d = -.13. In the next step, the confidence intervals of both effect sizes were calculated, [-.38 to .10] for the difference concerning the reference group of majority, and [-.37 to -.11] for the difference concerning the reference group of the minority (both confidence intervals with 95% confidence). Results revealed that both confidence intervals overlap, which indicates that there is no significant difference between (a) the difference between the control group and the reference group of the majority within the double-rating

method and (b) the control group and the reference group of minority within the double-rating method with reference to the setting of the simulated personnel selection procedure.

To determine whether the other-rating of the indirect questioning technique yields a different integrity score for either reference group (the majority or the minority) for the simulated setting, the difference of the other-rating obtained for the indirect questioning technique in both reference groups was compared with the control group using self-rating only for the simulated personnel selection setting. The effect sizes of Cohen's *d* were *d* = -.08 for the reference group of the majority and *d* = -.15 for the reference group of the minority. The confidence intervals of both effect sizes were [-.35 to .19] for the difference between the reference group of the majority and the control group and [-.43 to .12] for the difference between the reference group of the minority and the control group (both confidence intervals with 95% confidence). Results revealed that both confidence intervals overlap, which indicates that there is no significant difference between (a) the difference between the control group and the reference group of the majority and (b) the difference between the control group and the reference group of the minority in the indirect questioning technique for the simulated personnel selection setting.

To further control for the potential to combat faking in both rating methods, the question arose as to whether the double-rating method differs in either setting. Levene's Test regarding the homogeneity of variances was significant (p = .043), which means that homogeneity of variances could not be assumed. The Welch-ANOVA was thus chosen for its robustness (Field, 2013). Results revealed that the integrity test score of the double-rating method was larger in the simulated personnel selection setting than in the anonymous survey setting, Welch's F (1, 808.63) = 13.42, p < .001.

The indirect questioning method was also tested to determine whether it differs regarding either setting. Levene's Test of homogeneity of variances yielded no significant results (p = .758). Results revealed that the integrity test score of the indirect questioning method was larger for the simulated personnel selection setting than for the anonymous survey setting, F (1, 425) = 5.51; p < .05, partial η^2 = .013.

To control the power of the simulation of the personnel selection setting, both control groups were compared in both settings. Results revealed that the self-ratings only in both settings do not differ significantly, F (1, 205) = 1,18; p = .279, partial η^2 = .006.

6.4 DISCUSSION

Methods of preventing faking have been investigated by a plethora of researchers. In this thesis, both Hui's (2001) double-rating approach and the indirect questioning technique were used to investigate the prevention of faking on integrity tests in an experimental design. Moreover, both rating methods and their underlying mechanisms were further tested using different reference groups, an impression management scale, and control variables. In the last step of the discussion section, findings from an explorative analysis of both the double-rating method and the indirect questioning method within a simulated selection procedure are discussed. This explorative analysis serves as a first step in investigating the potential of both novel rating methods to combat faking on integrity tests for the application of these tests in personnel selection settings.

Reducing faking via the double-rating method and the indirect questioning method

Results from the analysis of variance revealed that the self-rating in the double-rating method was significantly different than in the self-rating only. This finding is in line with findings by Hui (2001) and Thomas et al. (2007). Moreover, results from the correlative approach revealed that the difference between the self- and the other-rating were significantly and positively correlated with the impression management scale. More specifically, this finding indicates that the more people evaluate themselves as being better than others, the more they tend to intentionally answer in a socially desirable way. Moreover, the more people evaluate themselves as being less better than others, the less they tend to intentionally answer in a socially desirable way. Taken together, these two results from the current study can be interpreted supporting Hui (2001) that the double-rating method can decrease socially desirable answering.

In the present study, the double-rating method led to reduced faking on integrity tests, which is an important finding as integrity tests are prone to faking (Alliger & Dwight, 2000; McFarland & Ryan, 2006). Moreover, the application of the double-rating method has additional advantages over many other methods that aim to reduce socially desirable responding: The double-rating method is simple, can be used on available tests, and prevents faking before it occurs instead of detecting it after it has occurred.

In contrast to the results of the double-rating method, the indirect questioning technique showed no significant difference between the other-rating and the self-rating only condition used in the control group. Nevertheless, the rating of the indirect questioning group showed the expected tendency, indicating that the integrity of the other-rating of the indirect questioning technique was lower than the self-rating of the control group. This finding contrasts with the significant difference of means found by Dalal and Hakel (2016), which might have been caused by differences in the explored construct: Dalal and Hakel (2016) investigated CWB, whereas the present study investigated integrity. Indirect questioning might increase more with negative behavior, such as CWB, than with desirable behavior, such as integrity. Another reason for the non-significant finding might be that there is a tendency to underestimate the effects of faking (Burns, Shoda, & Roebke, 2017). In greater detail, faking could have been underestimated if only effect sizes were considered and the prevalence of faking within participants was ignored: If only some participants fake, the effect size grows more strongly than if a high percentage of participants fake.

The question arose as to how large the influence of faking is on these integrity-related items. There may not have been a need for participants to fake answers or to strongly fake answers because the web-based survey in the current study was anonymous and did not involve any face-to-face communication with an advisor. If there had been a low level of faking, there would not have been a significant difference between the means of the groups for either method (double-rating or indirect questioning) and the control group (self-rating only) because participants would not have faked their answers to a significant extent.

The investigation into reducing faking on integrity tests via the double-rating method and the indirect questioning technique yielded contrary findings in the present study. Nevertheless, both previous studies and the present study have revealed promising findings suggesting that the double-rating method has the potential to combat faking. In contrast, indirect questioning was not successful at reducing faking within the current study. Regarding the heterogeneity of integrity tests and their items, more research must be conducted to further investigate both methods for integrity testing.

The reference groups of the majority and the minority

In his study, Hui (2001) assumed that if participants felt as though they belonged to a majority, they would more strongly admit to their own tendencies to behave undesirably or to have undesirable attitudes. Therefore, the feeling of being a part of the majority would lead participants to less socially desirable answering. This assumption was tested by comparing the reference group of the majority (people in general) and the minority (a single person) with the double-rating method. Results revealed that there were no differences in the integrity ratings between both reference groups of the double-rating method for either setting. This finding contradicts Hui's (2001) assumption that participants should feel comfortable and report CWB more readily if they believe that they are part of the majority.

An alternative explanation might be that participants tend to show a stronger contrast between their own and others' attitudes or behaviors. This means that participants might provide a bad rating for the integrity of others in order to cast these others in a bad light, thereby allowing themselves to appear to have greater integrity in contrast to the "bad" others, even if the participants' real integrity-relevant attitudes or behaviors might be worse than average. The pattern of the ratings in the double-rating method might provide support for this explanation: The mean of the other-ratings in the double-rating method is always worse than the mean of the self-ratings in the double-rating method. This pattern of higher self-rating scores than other-rating scores was also found by Hui (2001). Moreover, the difference between the self-rating and the other-rating displays a positive correlation with the impression management scale. This means that participants with a high tendency to fake (that is, those with high impression management scores) should show a stronger contrast between their self-rating and their other-rating. This finding additionally seems to support the assumption that participants rate others as having less integrity in order to strengthen the appearance of their own integrity. How participants act in real-life settings should then serve as a guideline for their own judgements of others' integrity.

In addition to the double-rating method, the reference groups of the majority and the minority were also added for the indirect questioning technique. Again, no difference between the reference groups was found for indirect questioning. This finding indicates that the type of reference group (majority vs. minority) makes no difference for the indirect questioning technique or for its potential to reduce socially desirable answering. The indirect questioning technique (which assumes that participants project their own attitudes or

behaviors in their other-rating) is stable with regard to a majority or minority reference group. The decisive element of the method of indirect questioning seems to be that respondents make judgements of others instead of judgements of themselves.

Analyses of control variables

Analyses of control variables were conducted with reference to both the double-rating and the indirect questioning method. Regarding indirect questioning, the control variable—which asked about the person or people to whom participants referred when answering the items—revealed that only two-thirds of the participants had themselves in mind when providing other-ratings in the indirect questioning method. This finding is important for indirect questioning because the method states that people project their own attitudes and behaviors when rating others (Dalal & Hakel, 2016). This assumption is controversial and seems to be partly falsified by the current results.

Due to similar results from polygraphs (Bergmann, et al., 1990; U.S. Congress, Office of Technology Assessment, 1983), concerns about the validation of the method of indirect questioning might arise. It is unclear what the remaining one-third of participants refer to when answering the other-rating in indirect questioning. They might unconsciously refer to themselves when answering the other-rating of the indirect questioning method. More research must be conducted to investigate the underlying mechanism of the indirect questioning technique.

In comparison, about 90 percent of participants claimed to have referred to themselves when responding to the experimental condition of self-rating only (the control group) and the self-rating of the double-rating method. Two conclusions are indicated by this finding: (a) There is no difference in response frequency for the control question (i.e., to whom participants referred when answering the integrity items) between the self-rating only and the self-rating of the double-rating method. (b) Not all participants claimed to have referred to themselves in the self-ratings, and some participants claimed to have referred to another person. This result is obviously due to reasons of misunderstanding or misreading. Another reason for this result might be that some participants only have referred to themselves unconsciously and therefore did not claim to have referred to themselves when answering. It would be interesting to examine participants who might have referred to themselves unconsciously when giving their integrity ratings in order to determine whether

they did so for all integrity items or whether their answers regarding the person or people to whom they have referred had varied between integrity items.

Interestingly, 20% of participants indicated that they had referred to themselves when answering the items for the other-rating of the double-rating method. This finding supports the assumption that participants might want to differentiate their own integrity by contrasting it with that of others. It would therefore be interesting to compare the difference between the self-ratings and the other-ratings in the double-rating method for participants who claimed to have referred to themselves in the other-rating with participants who claimed to have referred to others in the other-rating. If participants want to differentiate themselves from others, they might give a much higher integrity rating to themselves and a worse rating for the other-rating only when they claim to have referred to themselves in the other-rating. In addition, it remains possible that participants might unconsciously refer to themselves when providing an integrity rating of others. Given that this and further questions remain unanswered, more research is needed in this area, especially to validate the effects of projection on indirect questioning.

The identification variable investigated whether participants had identified with the person in the situation described in the items and revealed that most of the participants had identified or partly identified with the person in the described situation. This finding indicates that participants could more easily put themselves in the position of these integrity-related situations, and that the chosen situations matched closely with reality in companies.

Explorative analysis of simulated personnel selection setting

The first step of the explorative analysis investigated how realistic the simulation of the personnel selection setting appeared to participants. The results of one of the control variables revealed that participants could imagine the simulated application procedure. Moreover, the two control groups of both settings did not differ significantly. In contrast to this latter finding, other studies have revealed significant differences of integrity tests between anonymous and simulated personnel selection settings (Marcus, 2006). This finding casts some doubts as to whether the manipulation of the simulated personnel selection procedure was indicated strongly enough. Future research could give more-detailed instructions to participants in order to help them better imagine the simulated personnel

selection procedure. For example, the participants could be asked to write an application letter before beginning the integrity rating.

The main questions of whether both rating methods can reduce faking was also investigated within the explorative analysis of the simulated personnel selection setting. Neither the double-rating method nor the indirect questioning method differed from the control group within the simulated setting. This non-significant finding for the indirect questioning method in the simulated setting is in line with the non-significant finding of the indirect questioning method in the anonymous setting, which also showed no difference between the rating of the indirect questioning technique and the rating of the control group. In contrast to the significant finding of the double-rating method for the anonymous setting (i.e., the self-rating of the double-rating method and the self-rating of the control group differed), for the simulated setting the self-rating of the double-rating method showed no significant difference from the self-rating of the control group. This oppositional finding of the double-rating method for both settings might have been influenced by the present study (e.g., participants might not have been able to imagine the simulated selection setting well enough) or by the double-rating method itself, which might not work in a field with a high occurrence of faking, as is the case with personnel selection.

Interestingly, the group means of both rating methods (double-rating and indirect questioning) increased significantly in the simulated setting in comparison with the anonymous setting. In contrast, the group means of both control groups did not differ between the two settings. Therefore, both rating methods seem to be prone to faking in the simulated personnel selection setting. More research with simulated and real personnel selection settings is thus needed to investigate the double-rating method and the technique of indirect questioning and their potential to reduce faking within these settings.

Future research

The next step in investigating the potential of both novel rating methods to prevent faking on integrity tests involves using a within-design to indicate score changes in both methods. Comparing scores intra-individually under different experimental conditions can yield information on the differences in score changes and rating methods. All findings from this study could be tested and confirmed via the within-design. In so doing, results might

provide further insights into the double-rating method, the indirect questioning technique, or the degree to which they are prone to faking.

Another research approach could be used to vary experimental rating conditions. For example, participants could be instructed to do a maximum of faking ("fake-good" condition), to do a minimum of faking ("fake-bad" condition), or to be honest ("honest" response condition) with their integrity ratings for both rating methods and the control group. These three instructional sets represent classical techniques of research in the field of socially desirable responding. For example, for the fake-good condition, it would be interesting to examine for the double-rating method (a) how the difference between the self-rating of integrity and the other-rating of integrity develops and (b) how both integrity ratings (i.e., the self-rating and the other-rating) score in contrast to the self-rating of the control group. Moreover, the difference between rating conditions (e.g., the difference between fake-good and fake-bad or between the self-rating and the other-rating of the double-rating method for the fake-good condition) can provide additional insights into the rating methods.

Adding more control variables could further help to investigate the occurrence and degree of faking. For example, participants could be asked why they gave a lower integrity rating. A detailed interview with every participant about their cognitive process when answering might yield new insights into the double-rating method, the indirect questioning technique, or their underlying mechanisms.

Limitations

Although the present study goes beyond the bounds of previous research in several ways, it has several limitations. To begin, all variables were measured with self-reports, which go hand in hand with a common-method bias. Nevertheless, integrity tests are also self-reported in almost every case. Moreover, control variables (e.g., identification with the person described in the situation) helped to provide better insight into participants' ways of thinking when providing a rating.

Regarding the study's content, one limitation concerns the missing validation of the underlying mechanism of the double-rating method. It should be noted that this study represents a first attempt to transfer the double-rating method to integrity tests. Moreover, new conditions were tested for the double-rating method (i.e., reference group of the

majority vs. the minority, setting of anonymous survey vs. simulated personnel selection procedure). However, it is difficult to gather enough qualitative information to verify an underlying mechanism via validation data. Nevertheless, the method of testing difference between means is a widely used approach within research to verify a method's potential to reduce faking (e.g., Yu , 2008; McFarland & Ryan, 2000). In the present study, this method was used to test difference between group scores in both conditions (i.e., the double-rating method and the indirect questioning technique) and the control group (which was used to validate the potential of the double-rating method to reduce faking). Moreover, in order to strengthen the validation of the double-rating method, a second, correlative approach was used to validate the potential of the double-rating method to combat faking.

For the indirect questioning method (which is based on assumptions that have to be detected), validation data are also needed to verify the potential of this method to reduce faking (Dalal & Hakel, 2016). Hence, more research is needed to further investigate the psychological mechanisms step by step and to collect validation data on both the double-rating method and indirect questioning.

Another point of criticism might be that the survey was conducted online and included a web-based test format instead of face-to-face research using a paper-and-pencil test. Nevertheless, previous studies have revealed no difference in the occurrence of socially desirable responding on web-based settings or web-based test formats in contrast to the standard setting of a face-to-face situation or paper-and-pencil tests (Gnambs & Kaspar, 2017; Rossiter, 2009).

Conclusion

Faking is one of the most critical and most frequently observed challenges faced by integrity tests (Alliger & Dwight, 2000; Birkeland et al., 2006). In an effort to prevent faking, two rather novel rating methods—the double-rating method by Hui (2001) and the indirect questioning technique—were compared with a control group of self-rating only.

Findings imply that the use of the double-rating method might reduce faking because the integrity scores of the double-rating method were significantly lower than were those of the control group of self-rating only. In line with this finding, the difference between both ratings (the self-rating minus the other-rating) within the double-rating method revealed a

significant and positive correlation with impression management. This correlative finding strengthens the assumption that the double-rating method might be able to reduce faking.

The setting of simulated personnel selection yielded contrary findings: Scores in the double-rating method increased significantly in comparison with the double-rating method within the anonymous survey. Moreover, the double-rating method did not significantly differ from the control group that used self-ratings only. Nevertheless, the simulated personnel selection setting within the current study enabled an additional, explorative analysis to be conducted regarding faking on integrity tests. Further research should be performed to explore the double-rating method within selection settings.

In order to gain better insight into the underlying mechanism of the double-rating method, additional questions were analyzed. The question arose as to the validity of Hui's explanation of the method (i.e., that participants have the feeling of being part of a majority when providing the other-rating and thereby admit to more bad behaviors or attitudes). As a consequence, two reference groups were created: the reference groups of the majority and the minority. Results revealed that it was not essential for the mechanism of the double-rating method to refer to a majority of people because there was no difference between integrity ratings for either reference group (the majority or the minority).

With regard to the potential of the indirect questioning technique to reduce faking, results revealed no significant difference in means between indirect questioning and the self-rating only control group. Moreover, this method basically assumes that an individual's attitudes or behavior are projected in the other-rating (Fisher, 1993), but not all participants referred to themselves when responding to the other-rating in the indirect questioning method. With regard to the method of indirect questioning, some doubts remain about its effectiveness at reducing faking and about its validity.

Additional findings enabled greater insights into the indirect questioning method to be gained. For the reference group, no difference was found between when participants referred to a majority or to a minority within the other-rating of indirect questioning. Moreover, participants only partly referred to themselves when providing the other-rating in indirect questioning.

In light of these findings, the double-rating method seems to be a promising method for preventing faking on integrity tests. However, more data are needed for the double-rating method with regard to its validation and application (e.g., simulated vs. real

personnel selection procedures). In contrast, the indirect questioning method was not found to have any potential to reduce faking. Some questions are left regarding the mechanism of the indirect questioning method, namely whether participants have themselves in mind when answering. Validation data are urgently needed to verify this underlying assumption of the indirect questioning method.

CHAPTER 7 GENERAL DISCUSSION ON INTEGRITY

This dissertation further investigated *what integrity is* and *how integrity can be used*. In this chapter, all key findings of this dissertation are presented and finally discussed. Focus was placed on construct validity by reviewing integrity tests (Chapter 3) and expanding the nomological network of integrity (Chapters 4 and 5). Moreover, this dissertation focused on criterion validity by expanding both validity criteria and incremental validity (Chapters 4 and 5). Finally, it also examined criterion validity with the aim to reduce faking on integrity tests via two promising approaches related to the rating design of test items (Chapter 6). Broadly speaking, the main findings of this dissertation contribute to a better understanding of the construct of integrity, integrity tests, and the validity of these tests. Furthermore, in this chapter, the implications for future research and practice are discussed based on the main findings of this dissertation. In addition, the strengths and limitations of the present studies and the methods used are discussed, and avenues for further research are proposed. Finally, a conclusion summarizes the main findings of this dissertation and their implications.

7.1 DISCUSSION OF MAIN FINDINGS

Integrity is a poorly defined construct (Karren & Zacharias, 2007), and integrity tests have been developed to predict a broad range of CWB (Ones & Viswesvaran, 1998b). Four approaches were used in this dissertation to illuminate construct and criterion validity: (a) investigating the main correlates of integrity, (b) discovering integrity tests, (c) reducing faking, and (d) enlarging the criterion of integrity (see Figure 7). Moreover, these four approaches are closely related to the process of measurement. It is important to summarize the findings of both key questions, *what integrity is* and *how integrity is used*, while reflecting the perspectives of construct and criterion validity of integrity tests.

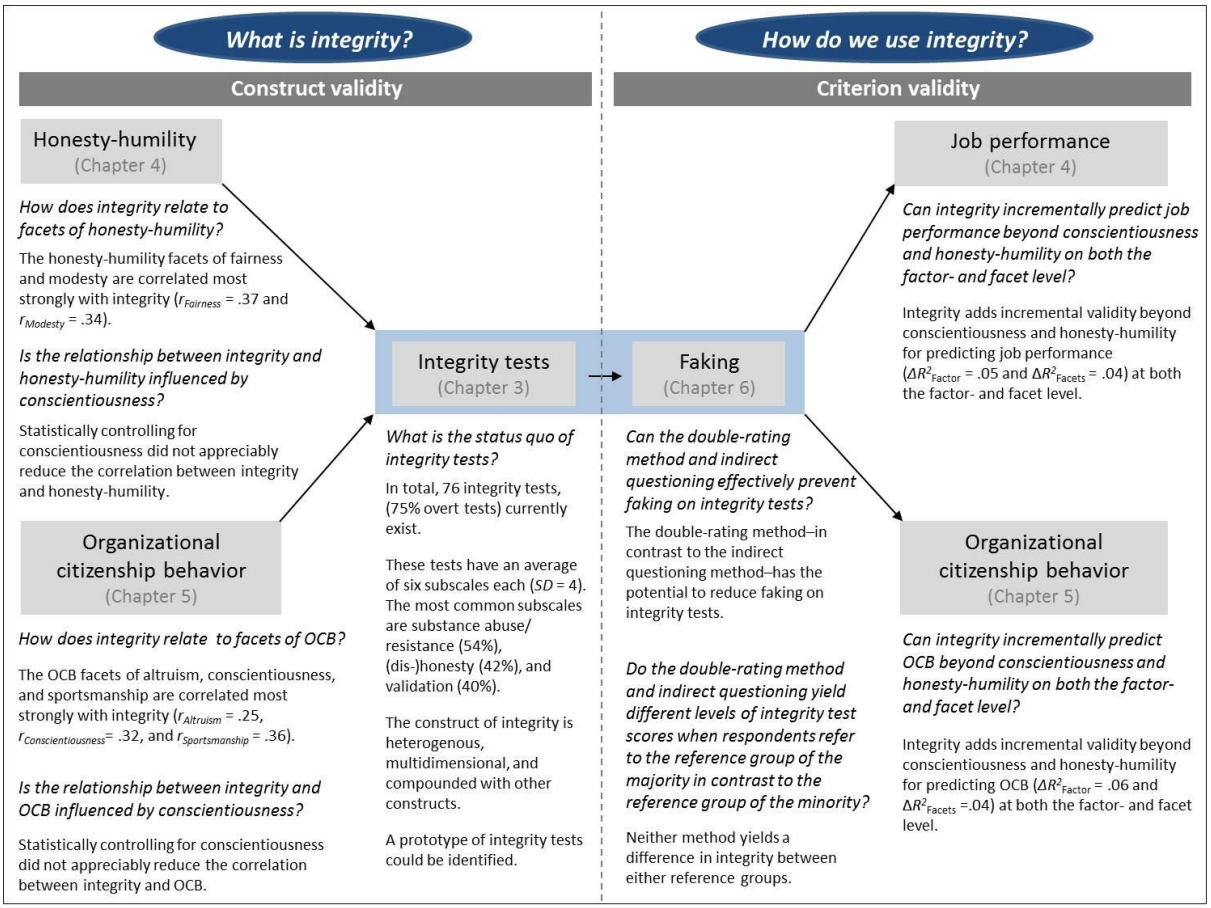

Figure 7. Overview of the focus of this dissertation, including its two key questions on construct and criterion validity and its main findings on the key questions from Chapter 1.3 about the central variables of the dissertation

What is integrity? – Main findings on the construct validity of integrity

This dissertation has yielded findings that contribute to the question of *what integrity is* in relationship to the construct validity of integrity tests. These findings concern (a) the overview and analyses of subscales as well as the construction and application of integrity tests (Chapter 3), (b) the relationship between integrity and honesty-humility at the factor- and facet level and the influence of conscientiousness on this relationship on the factor level (Chapter 4), and (c) the relationship between integrity and the behavior of OCB at the factor- and facet level, and the influence of conscientiousness on this relationship on the factor level (Chapter 5).

What is the status quo of integrity tests?

While integrity remains an ill-defined construct (Karren & Zacharias, 2007), integrity tests are used across the world. Studies and meta-analyses focus on the best-known integrity tests, yet nothing is known about the current number of integrity tests that exist worldwide. The most recent, albeit outdated report found 45 commercially used integrity tests in the United States (Ones & Viswesvaran, 1998b). In addition, only one overview could be found, which covered 43 integrity tests (O'Bannon et al., 1989). Both overviews were conducted more than 20 years ago.

Therefore, an up-to-date overview of current integrity tests was urgently needed. In comparison to the outdated numbers of 45 and 43 tests, this study has identified 76 integrity tests existing worldwide that meet the narrow criteria of an integrity test (CWB criterion and personnel selection setting) and demonstrated the predominance of overt integrity tests (75%). The findings reflect the vast number of integrity tests, which has constantly grown in recent decades. As a result and with reference to the diverse constructs of integrity, questions arise regarding the contents of integrity tests.

Overall, 50 integrity tests were reviewed regarding their subscales. There were strong quantitative variations in subscales, with an average of six subscales (*SD* = 4; maximum 26 subscales). This finding aligns with previous studies, which found several (most frequently four) factors of integrity by conducting factor analyses (O'Bannon et al., 1989; J. Hogan & Brinkmeyer, 1997; Wanek et al., 2003). The multidimensionality of the integrity construct can be supported by previous research and the findings of the present study.

The analyses of integrity test subscales further revealed that subscales vary greatly in content. This finding aligns with previous findings, which revealed that integrity tests measure different aspects (Nicol & Paunonen, 2002; Wanek et al., 2003). Through qualitative analysis, the most common subscales could be identified: In nearly half of the tests, the construct is implemented by the subscales of substance abuse/resistance (54%), (dis-)honesty (42%), and validation (40%). While (dis-)honesty is a very narrow facet of integrity, substance abuse/resistance forms part of CWB and therefore belongs to another construct of behavior. This finding is in line with other researchers who noticed construct confusion within integrity tests (Karren & Zacharias, 2007).

Given this huge variation in integrity tests, it was necessary to systematically compare tests with regard to their construction and their application. This systematic

analysis of integrity tests allows for a comparison of integrity tests at a glance for the first time. Previous overviews of integrity tests were either outdated (e.g., O'Bannon et al., 1989; Ones & Viswesvaran, 1998b), not systematic (e.g., Coyne & Bartram, 2002) or not comprehensive (e.g., Snyman et al., 1991).

In order to offer an up-to-date, systematic, multidimensional overview of integrity tests, sixteen such tests were selected for a detailed comparison with a focus on their construction and application. For this detailed analysis, integrity tests were selected according to three criteria: novelty, publicity, and variety. The criterion of publicity, which encompasses the frequency of test use and its citation in studies, meta-analyses, and additional literature, is important because the most of the main findings on integrity tests rely on only a handful of tests (Ones & Viswesvaran, 1998b). This criterion therefore offers a solid basis on which to select representative integrity tests.

The main findings of the systematic analysis concern the drawbacks of integrity tests. Integrity tests lack underlying models or theories as a result of the criterion-oriented process of test development. This finding is in line with results of previous studies (Karren & Zacharias, 2007). While some integrity tests benefit from a very large norm sample (e.g., 45,000 for the HPI-R), others have undersized norm samples (e.g., CBI) in comparison with the recommended norm sample size of 300 participants (Evers, 2001). In almost all cases, the construct of integrity is not defined in available test manuals.

Although the primary intention of an integrity test is to predict CBW, in almost cases, this definition was expanded to include predicting job performance, which is contrary to the classical definition of an integrity test. As a result, a clear distinction should be made between traditional integrity tests with a clear focus on predicting CWB and tests with a broad focus aimed at predicting a range of work-related behavior, including CWB, job performance, and OCB. The first kind of test (i.e., that with the narrow focus) could be renamed as the *compliance test*, and the second kind of test (i.e., that with the broad focus) could be named as the *integrity test*.

Integrity tests have been shown to include many subscales: based on the qualitative analysis of 50 integrity tests, they consist of an average of six subscales (*SD* = 4). A systematic comparison of 16 integrity tests found an average of eight subscales (*SD* = 6). With regard to the naming of these subscales, it became obvious that the construct of integrity is confused with other constructs (e.g., CWB, social desirability, burnout, mental dispositions, situational

characteristics). This finding is in line with previous findings (Nicol & Paunonen, 2002; Wanek et al., 2003). Furthermore, in addition to the large number of subscales, integrity tests consist of many items (*M* = 93 items; *SD* = 63 items). With regard to the kind of integrity test, overt integrity tests are predominant.

In summary, these main findings about integrity tests enable a prototype integrity test to be developed and recommendations for improving tests to be made (see Table 12). The prototype is important for integrity tests for two reasons: (a) Selecting a representative integrity test for research enables the results of integrity tests to be interpreted and compared and the findings to be generalized. (b) The choice of a suitable integrity test for personnel selection enables results about applicants' integrity to be put to best use.

Table 12

Overview of integrity tests' quality aspects, their status quo, and a prototype integrity test

Quality aspects of integrity tests	Status quo of integrity tests	Aspects of a prototype integrity test[a]
Aim	The aim is broad, with reference to identifying CWB and high-performing applicants and employees.	The classical aim of the test is to identify applicants' tendencies toward CWB.
Application	Most integrity tests have a web-based option, which is the form recommended by researchers.	The test is *(and should be)* available in a web-based form.
Content	The tests include many subscales, some of which form part of other constructs.	*The prototype should consist of one subscale that includes a clear and narrow focus on CWB-related element.*
Duration	It takes a long time to complete the tests as a result of the many included items.	*The test should be of a short duration with fewer items (maximum 15 minutes).*
Format	There is a predominance of overt tests.	The test is an overt test.

Norm sample	The size of the norm sample varies and is missing in some cases.	*The test should refer to a norm sample of at least 300 working participants.*
Theory	An underlying theory is missing in almost all integrity tests.	The test is based on a criterion-focused construction.

Note: The quality aspects of integrity tests are presented in alphabetical order. a = The aspects of the prototype in italics represent recommendations adapted from different integrity test manuals.

How does integrity relate to facets of honesty-humility? Is the relationship between integrity and honesty-humility influenced by conscientiousness?

As a result of its multidimensional characteristic (e.g., Wanek et al., 2003), integrity is influenced by different personality traits. Conscientiousness was shown to be the most stable and important personality variable relating to integrity alongside agreeableness and emotional stability (Ones, 1993; Marcus et al., 1997). In addition to conscientiousness, honesty-humility was also strongly correlated with integrity (Marcus, Lee, & Ashton, 2007; K. Lee, Ashton, & De Vries, 2005), a finding confirmed by this dissertation (r = .45). This finding indicates that honesty-humility has a high degree of identical conceptual connection to integrity and further supports the congruence of both constructs (e.g., Laginess, 2016). As a result, honesty-humility has shown to be one of the key related personality constructs in the nomological network of integrity. However, two gaps exist in past research: first, the strength and direction of the relationship between integrity and the facets of honesty-humility remain unclear, and second, it remains unclear whether conscientiousness can diminish the relationship between integrity and honesty-humility.

With regard to the first question, previous studies have concentrated on the relationship between honesty-humility at the factor level and integrity (K. Lee et al., 2008), or on the relationship between honesty-humility at the factor level and CWB (K. Lee, Ashton, & De Vries, 2005). This dissertation provided the missing focus on honesty-humility at the facet level and integrity (K. Lee, Ashton, & De Vries, 2005), revealing that every facet of honesty-humility is significantly and positively correlated with integrity. The strongest facets are fairness (r = .37) and modesty (r = .34). The facet of fairness assesses a tendency to avoid fraud and corruption. People with low scores are willing to gain by cheating or stealing, whereas people with high scores are unwilling to act unfairly to gain from other individuals

or society. Fairness, which also formed—in some definitions—part of the concept of integrity (e.g., T. E. Becker, 2005; Barnard et al., 2008), is important in terms of obeying rules and acting genuine toward others. Both following rules and socially fair-minded behavior are important aspects of integrity, especially in terms of the primary criterion of CWB. This conclusion is supported by a study that found the fairness facet of honesty-humility to be the most important predictor of delinquency (R. E. De Vries & Van Gelder, 2013). Moreover, the fairness facet even explained almost as much variance in workplace deviance as all six HEXACO domains combined (Pletzer et al., 2020). These previous findings are in line with the present finding. As a result, the fairness facet should be included in the definition of integrity and represents a key facet of the construct of integrity and its nomological network.

The facet of modesty presents integrity in a different light: Modesty refers to the personal characteristic of being moderate and unassuming. People with low scores consider themselves to be superior and entitled to privileges that others do not have, whereas people with high scores view themselves as ordinary people with no deservingness of special treatment. As a result, egoism and selfishness seem contrary to integrity. To my knowledge, no study has yet investigated the relationship between integrity and modesty. Focusing on the concept of modesty, which is being humble instead of intent on one's own advantages, modesty is clearly linked to integrity. The correlative relationship between integrity and modesty reveals that modesty is a part of integrity and its nomological network.

With regard to the second question, previous studies have only controlled for conscientiousness in the relationship between integrity and job performance (Murphy & Lee, 1994a) or have controlled for honesty-humility in the relationship between integrity and CWB (Laginess, 2016). The current findings of this dissertation revealed that the relationship between integrity and honesty-humility is not weakened by partialling out conscientiousness: Honesty-humility does not overlap with conscientiousness and completes the concept of integrity; it has an additional and individual conceptual connection to integrity. Although conscientiousness is one of the most important empirical correlates of integrity (Murphy & Lee, 1994b; Ones et al., 1993) and similarities exist in items and subscales for both constructs (J. Hogan & Hogan, 1989), honesty-humility was found to be another important component of the construct of integrity, independent of the overlap between conscientiousness and integrity.

How does integrity relate to facets of OCB? Is the relationship between integrity and OCB influenced by conscientiousness?

In contrast to the vast amount of research on integrity and job performance or CWB, OCB is a long-neglected albeit noteworthy correlate of integrity. The behavior of OCB is strongly linked to integrity in two ways: (a) OCB is associated with CWB, which is the primary criterion of integrity tests; thus, OCB and CWB are theoretically and empirically linked (Borman & Motowidlo, 1993; Dalal, 2005; Sackett et al., 2006). (b) Integrity and OCB appear to overlap conceptionally (Casillas et al., 2009; Hertel et al., 2000). The empirical relationship between integrity and OCB has been investigated (Van Iddekinge et al., 2012a). In one study conducted by Casillas et al. (2009), the relationship between integrity and OCB might have been overestimated because the study struggled with some methodological problems, particularly because the measures of OCB and integrity were confused with other constructs. As a result, it appeared necessary to reinvestigate the empirical relationship between integrity and OCB.

To empirically demonstrate the relationship between integrity and OCB, this dissertation revealed a relationship of r = .37 for both variables, which is a moderate effect with regard to the number of participants. Casillas et al. (2009) found a considerably higher relationship between their integrity test and OCB (r = .70). However, in contrast to the present study, the researchers had a restricted focus on one-sided facets of their integrity test (general work attitudes and risk reduction) which do not represent the core construct of integrity. Moreover, the construct of integrity used in their integrity test was confused with job performance and OCB. Therefore, the empirical results of Casillas et al. (2009) must be interpreted with caution. Considering past research, two gaps can be identified: First, the strength and direction of the relationship between integrity and the facets of OCB remain unclear, and second, it remains unclear whether conscientiousness can diminish the relationship between integrity and OCB.

To provide an answer to the first research gap, this dissertation takes a more detailed look into integrity and facets of OCB. The OCB facets of sportsmanship (r = .36), conscientiousness (r = .32), and altruism (r = .24) are significantly related to integrity in contrast to the OCB facet of civic virtue (r = .07, n.s.). All three significantly related facets of OCB are linked to integrity through their content.

The facet of sportsmanship reflects the ability to tolerate adversity, stress, and strains in the working environment. This definition of the facet sportsmanship is important because former studies have revealed that stress in the working environment has the potential to boost CWB (De Clercq, Haq, & Azeem, 2019; Striler, Shoss, & Jex, 2020). Moreover, Spector and Fox (2002) found that positive emotions lead to OCB, whereas negative emotions (e.g., anger or anxiety) lead to CWB. As a result, OCB along with its antecedent of positive emotions is a counterpart of CWB and its negative emotions, whereas the facet of sportsmanship is the essential component.

The facet of conscientiousness refers to diligence of employees, such as thoroughly preparing work. Conscientiousness as a facet of OCB is in line with conscientiousness as a personality trait; thus, the strong relationship between the conscientiousness facet and integrity is clearly a key component of integrity.

The facet of altruism includes helpful behavior, such as helping a colleague with work-related problems and reflecting positive employee characteristics. Altruistic people focus on those around them in order to help them. In contrast, people with no integrity seem to focus on their own needs and wishes. Moreover, given that integrity tests predict a wide range of CWB including not to consider the interests of other people (Bennett & Robinson, 2000), this conception of integrity fits with the OCB facet of altruism.

Taken together, these three facets of OCB and their relationship with integrity indicate their strong connection to the construct of integrity. These facets must be considered in the definition and nomological network of integrity. Moreover, the overall construct of OCB showed an empirical and relevant connection to the construct of integrity, and the construct of OCB therefore must also be taken into account when defining integrity and conducting research on integrity.

To address the second research gap, the relationship between integrity and OCB was controlled for conscientiousness. To do so, the strength of the relationship between integrity and conscientiousness had to be considered: The personality trait of conscientiousness was found to have a relationship of r = .32 with integrity. Conscientiousness, which is one of the most important correlates of integrity, showed a similar relationship with integrity in previous studies of r = .36 (Murphy & Lee, 1994a), and r = .42 (Ones et al, 1993). Moreover, in the present study, the relationship between integrity and OCB was found to be significant and to have comparable strength. This relationship between integrity and OCB was

demonstrated to be stable as it retains the same strength, even when conscientiousness was partialled out of the relationship between integrity and OCB. This finding is in line with Murphy and Lee (1994a), who found that controlling for conscientiousness had only a slight effect on the correlation between integrity and performance. As a result, the current finding of this dissertation indicates that OCB—including the facet of conscientiousness—is relevant to integrity and includes elements beyond conscientiousness.

Based on the findings of this dissertation concerning integrity and OCB, it can be stated that OCB is an important component of the nomological network of integrity and should urgently be considered within the definition of the construct of integrity. OCB, which is shown through positive behavior, such as helping colleagues and improving organizational tasks and processes, fits perfectly with integrity, which includes resisting CWB, obeying the rules and other elements of acting as a good employee.

In line with the findings on integrity and OCB, many concepts of OCB include the dimension of compliance (Borman & Motowidlo, 1993; Graham, 1991; C. A. Smith et al., 1983; Van Scotter & Motowidlo, 1996), also called generalized compliance (C. A. Smith et al., 1983), organizational obedience (Graham, 1991), or job dedication (Van Scotter & Motowidlo, 1996). Altogether, these aspects of compliance describe what a "good employee" should do or not do and the respect that should be afforded to rules and instructions, such as being on time or working without wasting time. Interestingly, this OCB dimension of compliance was found to correlate most strongly with the personality variable of conscientiousness, which strongly relates to integrity (Organ & Ryan, 1995). Moreover and in line with the CWB concept of CWB-I and CWB-O (Bennett & Robinson, 2000; Robinson & Bennett, 1995), Williams and Anderson (1991) devised the concept of OCB-O, which refers to behavior that benefits the organization in general, such as giving notice when one is ill.

CWB, the primary criterion of integrity tests, is defined as intentionally damaging the legitimate interests of the company (Nerdinger, 2008; Marcus & Schuler, 2004) and involves breaking rules (Marcus, 2000). Integrity tests focus on rule-following behavior in contrast to the rule breaking behavior of CWB. Thus, rule-following behavior could be described as a fundamental component of integrity. Integrity in this narrow sense of compliance or rule-following is defined as a behavioral construct. In some integrity tests, rule-following or compliance appear to be a subdimension of integrity (e.g., ARP) and form a part of the

definition of integrity (Murphy, 2000). In addition, some researchers emphasize a focus on narrowly defined dimensions of integrity (Van Iddekinge et al., 2005).

Summary of the findings on construct validity

The findings of this dissertation help to achieve a more precise focus on the construct of integrity, its definition, and its nomological network. Moreover, the findings support the construct validity of integrity because further and relevant correlates of integrity could be empirically identified.

Based on these findings about the construct of integrity, the definition of integrity given in this dissertation should be expanded. This revised definition of integrity, which may prompt a universal definition, is as follows:

Integrity is a multidimensional compound construct based on key personality traits (i.e., honesty-humility, conscientiousness, self-control), personality facets (i.e., fairness, modesty), attitudes (i.e., fairness), work-related behavior (i.e., OCB), and facets of work-related behavior (i.e., sportsmanship, conscientiousness, altruism). Moreover, by definition, integrity includes following laws, organizational rules, and individual rules, which are based on individual attitudes of what is right and what is wrong. Therefore, one of integrity's key components is compliance. Moreover, integrity is clearly related to work and organization, with individual attitudes and behavior directed at behaving correctly toward others, supporting others, and not insisting on acting to one's own advantage.

How do we use integrity? – Main findings on the criterion validity of integrity

The main findings of this dissertation revolve around the question of *how we use integrity*, and relate to the criterion validity of integrity tests. These findings concern (a) the incremental validity of integrity for job performance beyond two strong correlates of integrity—the traits of conscientiousness and honesty-humility—on both the factor- and facet level (Chapter 4), (b) the incremental validity of integrity for OCB beyond two strong correlates of integrity, which are the traits of conscientiousness and honesty-humility, again on both the factor- and facet level (Chapter 5), and (c) two promising rating methods for reducing faking (Chapter 6).

Can integrity incrementally predict job performance and OCB beyond honesty-humility and conscientiousness on both the factor- and facet level?

Integrity tests are a good predictor of job performance (Ones, 1993; Schmidt & Hunter, 1998). Moreover, a meta-analysis revealed that integrity tests have incremental validity for predicting job performance beyond a measure of general mental ability (Schmidt & Hunter, 1998). Regarding the missing correlation between integrity and cognitive ability (Ones et al., 1993), incremental validity could be easily found in this combination of measures. With regard to other predictors aside from general mental ability, most studies have not investigated integrity as a second predictor in addition to other integrity-related personality factors or facets for incrementally predicting job performance.

One of the strongest correlates of integrity is the personality trait of conscientiousness (Ones et al., 1993), which has also been shown to be one of the strongest predictors of job performance (Barrick & Mount, 1991). Another important correlate of integrity is the personality trait of honesty-humility, as has been demonstrated in previous studies (K. Lee, Ashton, & De Vries, 2005; Marcus, Lee, & Ashton, 2007) and additionally in the current dissertation. Honesty-humility has been shown to incrementally predict performance criteria beyond the Big Five factors, including conscientiousness (Catano et al., 2018; A. De Vries et al., 2011; M. K. Johnson et al., 2011). Both conscientiousness and honesty-humility have been shown to explain the incremental validity of performance criteria (A. De Vries et al., 2011; Dudley et al., 2006).

Although conscientiousness and honesty-humility are two strong correlates of integrity and both traits also strongly and incrementally predict job performance, in the present study, integrity added validity to the prediction of job performance beyond the factor and facets of both conscientiousness and honesty-humility ($\Delta R^2 = .05$, $p < .001$ for the factor level and $\Delta R^2 = .04$, $p < .001$ for the facet level). Therefore, this finding is novel because no study before has provided evidence of the incremental validity of integrity tests over these two most relevant correlates of integrity—namely conscientiousness and honesty-humility, which were simultaneously included in this study. Moreover, the current finding broadens the power of integrity tests to incrementally predict work-related behavior beyond other personality traits. This finding is important because such studies on the incremental power of integrity tests over other personality traits are rare (Marcus, Te Nijenhuis, et al., 2016).

In this dissertation, not only job performance but also OCB were incrementally predicted by integrity tests. The findings also revealed that integrity tests incrementally predict OCB and its facets beyond the factor- and facet level of conscientiousness and honesty-humility (ΔR^2 = .06, p < .001 for the factor level and ΔR^2 =.04, p < .001 for the facet level).

This finding is impressive because in contrast to CWB, OCB is an important work-related behavior that has thus far often been empirically neglected in the context of integrity tests. Although OCB is both theoretically (Borman & Motowidlo, 1993) and empirically (Dalal, 2005) strongly related to CWB, which is the primary criterion of integrity tests, only a relatively small number of studies have investigated the validity of integrity tests for predicting OCB (Casillas et al., 2009; Van Iddekinge et al., 2012a). Therefore, this finding is important because it sheds light on the dark environment surrounding the power of integrity to incrementally predict OCB on both its factor- and facet level. Moreover, this finding fills the gap that was left by previous studies that investigated the incremental power of integrity for predicting job performance or CWB (e.g., Schmidt & Hunter, 1998; Marcus et al., 2013).

The findings of the current study on the incremental gain in validity of integrity tests for predicting both job performance and OCB must be compared with previous and rare research studies that investigated the potential of incremental validity to predict work-related behavior via different combinations of predictors. To make this comparison, the current findings are discussed in light of two recent studies that focused on incremental validity in combination with integrity tests (Catano et al., 2018; Y. Lee et al., 2019).

(a) One recent meta-analysis investigated the incremental validity of honesty-humility beyond the Big Five traits, general mental ability, and integrity tests (these three variables were included in the first step of the hierarchical regression analysis) for predicting three different forms of organizational behavior (i.e., CWB, OCB, and task performance) (Y. Lee et al., 2019). Results indicated that honesty-humility has incremental

validity for CWB but adds no validity for task performance or OCB.[9] As general mental ability did not correlate with integrity and honesty-humility, the Big Five personality dimensions and integrity outperformed honesty-humility.

This finding is in line with those of the current study, which revealed that integrity adds validity beyond honesty-humility in predicting job performance and even OCB. In conclusion, integrity seems to be a better predictor of job performance and OCB than honesty-humility because it has the power to explain more variance in the criterion when making predictions than would honesty-humility. In addition, from the perspective of a practitioner, the predictors used in the current study demonstrates parsimony in contrast to the meta-analytic finding of Y. Lee et al. (2019), which additionally included the predictors of openness, extraversion, agreeableness, and emotional stability as well as a general mental ability test. In a personnel selection procedure, these additional predictors would be highly expensive in terms of both time and costs.

(b) Another recent study deserves to be compared with the current finding on the incremental validity of integrity tests for predicting OCB. This previous study found that integrity tests can incrementally predict CWB beyond the Big Five traits with a slight increase in predictive validity of $\Delta R^2 = .02$ (Catano et al., 2018). As CWB and OCB are empirically and conceptionally related (e.g., Borman & Motowidlo, 1993; Dalal, 2005) and appear to be two components along one continuum (e.g., Bennett & Stamper, 2001), this previous result fits with that of the current study. Again, a drawback of the used integrity test should be considered: Because integrity is a conglomerate of agreeableness, conscientiousness, and emotional stability (Ones, 1993), the integrity measure was constructed with carefully selected items from a Big Five measure. Therefore, some of the items cannot reflect the key core of integrity, and this missing key core might have influenced the strength of the results.

Despite this methodical drawback, one previous study supported the finding of Catano et al. (2018). Marcus et al. (2013) found that integrity tests have incremental

[9] In their study, the authors included delinquency measures as integrity tests in their meta-analytic analyses in order to obtain a more stable dataset (Y. Lee et al., 2019). However, it is important to note that delinquency measures are not integrity tests and must therefore be distinguished from traditional integrity tests. The researchers controlled for the influence of the kind of test via a moderator analysis and found a stronger correlation between honesty-humility and integrity for delinquency measures in contrast to both kinds of traditional integrity tests. This methodological issue might have led to the incremental validity of honesty-humility over integrity for predicting CWB.

predictive power for CWB over the HEXACO model. As the HEXACO model includes the factors of both conscientiousness and honesty-humility, this previous finding also strongly supports the current finding of the ability of integrity tests to predict work-related behavior. Moreover, the small gain in validity for predicting job performance and OCB found in the present study is in line with a study by Marcus et al. (2013), who found incremental predictive power of integrity tests for CWB over the HEXACO model of about the same magnitude (ΔR^2 =.034 and .059) (Marcus et al., 2013).

When comparing the factor- and facet level of predictors in the present study, the incremental validity of integrity tests for predicting job performance and OCB beyond relevant facets of conscientiousness and honesty-humility was found to be equally as strong as the incremental validity of integrity tests for predicting job performance and OCB beyond the factor level of both personality traits. In greater detail, no single facet of honesty-humility significantly explained variance for predicting job performance beyond the facets of conscientiousness. The current finding does not completely fit with the previous result, which found that both conscientiousness and honesty-humility explained the incremental validity of performance criteria at the facet level (A. De Vries et al., 2011; Dudley et al., 2006).

When predicting job performance in the first step of the hierarchical regression analysis in the present study, the conscientiousness facet of diligence had the greatest significant gain in predicting job performance (β = 0.22, $p < .001$), and the conscientiousness facet of prudence also significantly explained variance (β = 0.18, $p < .01$). This result was in line with the results of A. De Vries et al. (2011), who found that the conscientiousness facet of diligence produced the greatest significant gain in predicting academic performance.

With regard to the prediction of OCB in the current study, the honesty-humility facets of fairness (β = 0.25, $p < .001$) and greed avoidance (β = -0.14, $p < .05$) explained variance for predicting OCB over the significant conscientiousness facet of diligence (β = 0.25, $p < .001$). This finding is in line with previous findings, which revealed that the fairness facet, in particular, has power to add validity for predicting CWB–which is also contextual work-related behavior like OCB (e.g., R. E. De Vries & Van Gelder, 2013; Pletzer et al., 2020)–and moreover, that the fairness facet has incremental predictive power for predicting CWB (e.g., O'Neill et al., 2011).

To conclude, in comparing the factor- and facet models while considering both model fit and parsimony, the factor models were found to be superior to the facet models.

In summary, the findings of the incremental prediction by integrity tests of job performance and OCB beyond conscientiousness and honesty-humility remarkably expanded the criterion and incremental predictive validity power of integrity tests. These findings are especially meaningful in the context of the application of integrity tests within selection procedures. Integrity tests have been shown to add validity beyond other common personality constructs in predicting both job performance and OCB. Therefore, the use of an integrity test can be recommended for situations, in which predictive validity and a gain of validity are essential.

Can the double-rating method and indirect questioning effectively prevent faking on integrity tests? Do the double-rating method and indirect questioning yield different levels of integrity test scores when respondents refer to the reference group of the majority in contrast to the reference group of the minority?

Faking or impression management is still a significant and unresolved problem within integrity tests (Alliger & Dwight, 2000; Birkeland et al., 2006), which are highly prone to faking (McFarland & Ryan, 2006). Although many approaches and methods exist to avoid faking, no method has yet been found that effectively and easily does so (Burns & Christiansen, 2011). As a consequence, new approaches to reduce faking are urgently needed in general, and particularly for integrity tests (Goffin & Boyd, 2009; Karren & Zacharias, 2007).

Two promising types of rating methods, the double-rating method developed by Hui (2001) and the indirect questioning technique (Fisher, 1993), were tested to avoid faking on integrity tests. For the method of indirect questioning, no significant difference was found between indirect questioning (ratings only refer to the perspective of others) and a control group (ratings only refer to the individual's own perspective). This finding indicates that the indirect questioning method seems to have no potential to prevent faking on integrity tests. In contrast to this finding, previous studies found a potential for the technique to reduce faking (Fisher, 1993; Dalal & Hakel, 2016). The difference in results may be due to different variables that were used in these studies: While Dalal and Hakel (2016) used indirect questioning to reduce faking on a CWB measure, indirect questioning was used in the

present study to reduce faking on integrity tests. Both constructs appear to represent opposite poles on a positive–negative continuum of work-related behavioral variables (e.g., Bennett & Stamper, 2001). Moreover, some researchers stated that OCB and CWB represent two distinct constructs (Sackett, Berry, Wiemann, & Laczo, 2006). This dissimilarity of both organizational behavioral outcomes may imply that the two constructs have different relationships to faking and may therefore yield different findings. In addition, validation data are missing for the indirect questioning method in order to verify the potential of this method to reduce faking.

For the double-rating method, a different pattern of results was found. In contrast to the indirect questioning technique, the method comprises two ratings: the other-rating—referring to the rating perspective of others and the self-rating—referring to the individual's own rating perspective. To test the effectiveness of the double-rating method in reducing faking, two statistical tests were conducted: First, the difference between the self-rating in the double-rating method and a control group using self-rating only was tested. Second, the correlation of the difference between ratings of the double-rating method (i.e., self-rating minus other-rating) and an impression management scale was tested. The double-rating method yielded significant results for both statistical approaches, which indicates that this method has the potential to prevent faking on integrity tests (mean difference: $F(1, 515) = 3.87$, $p < .05$, partial $\eta^2 = .007$; correlative result: $r = .41$, $p < .001$). With regard to the small albeit significant result (and additionally to the non-significant result of the indirect questioning method in this study) that the double-rating method can reduce faking, effects of faking might have been underestimated due to a small proportion of faking participants (Burns et al., 2017).

This significant finding is important for integrity tests because it offers a novel application of this method to reducing faking in integrity tests. The new method has three primary advantages over other common methods that detect faking: First, it prevents faking before it occurs. Second, it is easy to apply in a test. Third, it can be used for integrity tests that already exist. Moreover, further information on faking can be obtained from the difference between the rating perspectives because systematically evaluating the gap between the perspectives may lead to a better understanding of the underlying mechanisms of faking.

Regarding the underlying mechanism of both the double-rating method and the indirect questioning method, the question arose as to whether a majority is needed for a reference group. Findings revealed that there is no significant difference for either method when the reference group refers to a majority of people in contrast to a single person. Regarding the double-rating method, this finding contrasts with Hui's (2001) suggestion and further contributes to the understanding of these rating methods. Moreover, it also helps to provide a more diverse range of applications for these methods in both research and practice.

In contrast to the significant findings of the double-rating method within the anonymous survey setting, a first exploratory approach to test the double-rating method within a simulated selection setting revealed contrary, non-significant results on the power to reduce faking. Nevertheless, this dissertation offered a first insight into the application of this novel technique of double-rating in integrity tests. It adopted novel method from another field of psychology to fill a research gap by finding a method to effectively reduce faking on integrity tests. Nevertheless, further research is needed to investigate the double-rating method and its validity in reducing faking.

In summary, the potential of the double-rating method to avoid faking is important and should be considered in integrity tests in further studies. Moreover, by reducing faking on these tests, the error of validity can be reduced and, as a result, the criterion validity can be enhanced.

Summary of the findings on criterion validity

The findings in this dissertation help to expand the focus on the criterion of integrity and especially on incremental validity. Integrity showed to have incremental power for predicting different work-related behaviors and outperformed other personality predictors on the factor- and facet level, even when these predictors are closely related to integrity. Moreover, the double-rating method might be a promising approach for strengthening the criterion validity of integrity.

7.2 IMPLICATIONS FOR RESEARCH AND PRACTICE

Integrity is a topic of current interest within research and practice: With reference to research, unresolved issues remain in this young field of research. With reference to practice, integrity tests have proved to be important in successfully predicting work-related behavior. The findings of this doctoral thesis have supported the research on integrity and raised implications for both integrity research and the practice of integrity tests as presented in the following section.

Implications for research

Integrity research still disagrees on the construct of integrity (Van Iddekinge, et al., 2012b). Much research has been conducted the nature of integrity and define the construct of integrity (Sackett & Schmitt, 2012; Sackett & Wanek, 1996), but these questions have been answered only incompletely and research gaps still exist (Karren & Zacharias, 2007).

The data in this dissertation focus on the main correlates of integrity and, in particular, on facets of these correlates in order to provide a clear understanding of integrity. Previous studies have frequently concentrated on the factor level instead of the facet level and on personality traits instead of behavior surrounding the nomological network of integrity. In contrast, this dissertation investigated both the facet level of the relationship between integrity and honesty-humility as well as OCB, and controlled for the relationship between integrity and honesty-humility as well as OCB for conscientiousness.

The honesty-humility facets of fairness and modesty were revealed to be important correlates of integrity, and both must be considered when investigating the construct of integrity. With regard to behavior, OCB and its facets of altruism, conscientiousness, and sportsmanship were shown to be important behavioral correlates of integrity. Moreover, despite controlling for conscientiousness, the relationship between integrity and both honesty-humility and OCB remained equally strong. As a result and in summary, these factors and facets are important elements in the construct of integrity and add to its definition.

When focusing on both the construct of integrity and the validity criteria of integrity tests, it can be beneficial to concentrate more on narrow facets. This approach concerns the bandwidth–fidelity dilemma (Cronbach & Gleser, 1957), which is a well-known hypothesis in the area of psychological testing (Salgado, 2017) and especially in assessing personality for

personnel selection purposes (Ones & Viswesvaran, 1996). This hypothesis refers to two aspects of a measure that must be considered when choosing a diagnostic test: its bandwidth and fidelity. In greater detail, researchers (and practitioners) have to determine whether they need a careful measurement of either one narrowly defined variable or a broad range of variables. While broad measures predict broad criteria very well, narrow measures predict narrow criteria very well (Cronbach & Gleser, 1957).

Previous studies have found that traditional integrity tests are measures with low bandwidth and high fidelity (e.g., Murphy, 2000), including a construct that is even broader and more complex than other personality traits, such as the Big Five traits (Ones & Viswesvaran, 1996). As a result, these tests best predict broad criteria, such as CWB (e.g., Ones et al., 1993; Van Iddekinge et al., 2012a). Keeping the vast variety of CWB (as the primary prediction criterion of integrity tests) (e.g., Gruys & Sackett, 2003) in mind, it is essential to explore and use narrow facets of integrity in narrowly constructed integrity tests. To be specific, in the current studies, two personality-based facets of honesty-humility–namely fairness and modesty–and three behavioral facets of OCB–named altruism, conscientiousness, and sportsmanship–were found to be substantially relevant for the construct of integrity. These facets might assist in determining a first approach to integrity tests with a focus on a narrow behavioral facet, such as helping behavior. Such narrow behavioral facets might be relevant for a special setting, for example, when finding a suitable occupational safety officer for a company by using a narrow integrity measure with a focus on complying with safety norms and avoiding safety risks.

When connecting these results on narrow facets with previous attempts to define integrity or integrity tests, the aspect of compliance can also be seen to form part of many OCB concepts and seems to be a key aspect of a narrower construct of integrity. Compliance, which can be described as rule-following behavior, reflects the notion that integrity has no tendency to CWB. Thus, compliance is also a key predictor of CWB, in line with the traditional use of integrity tests to predict only CWB rather than more diverse criteria (e.g., job performance). Renaming integrity tests as *compliance tests* and renaming the underlying construct of integrity as *compliance* would have many advantages: (a) It would emphasize that the element of compliance is the main component of the construct. As a result, a general, widely accepted definition of integrity may follow. (b) It would help to retain the original, narrow focus of this kind of test and to support a narrow focus on the underlying

construct, which is urgently needed. Furthermore, the view of compliance as a narrow concept of integrity may provide new impetus to tackling the problem of the vast variation in the content of integrity-test subscales and items. (c) There would no longer be any confusion over the term integrity when used in other disciplines (e.g., philosophy, criminology, business management) and with a different meaning (e.g., appearing as an individual moral virtue or a company virtue) (Audi & Murphy, 2006).

The challenges presented by diverse integrity tests might be also reduced by creating a prototype integrity test that includes the most commonly found characteristics of this kind of test. Such a prototype would enable researchers to interpret and compare results between different integrity tests. Moreover, using a prototype integrity test better supports the generalization of results. Regarding the wide variety of integrity tests, a prototype could support a clear concept of integrity and a more straightforward type of integrity test.

Additional implications arise from different parts of the review of integrity tests: (a) The collection of currently existing integrity tests helps researchers to obtain an overview of existing tests and to choose a suitable test. (b) The quantitative and qualitative analyses of the subscales further support researchers in their test choice. (c) The systematic comparison of selected integrity tests gives insight to researchers about particular features of the tests and the strengths and drawbacks of each test.

The investigation of a novel method to reduce faking on integrity tests also provides benefits for future research studies. The method yielded new insights into the mechanisms and occurrence of faking. Concerning the mechanism of faking, the rating of other people sheds new light on the research field of faking. Concerning the occurrence of faking, the use of the double-rating method to prevent faking within research settings may help to reduce measurement errors, thus increasing the validity of research results regarding integrity and integrity tests.

Implications for practice

Integrity tests predict the primary criteria of CWB and job performance very well (Ones et al. 1993; Van Iddekinge et al., 2012a). Indeed, integrity tests are superior to many other measures in predicting job performance (Schmidt & Hunter, 1998). These former findings are important in the primary application of integrity tests within personnel selection settings, where companies make significant investment to select new employees with high

levels of job performance. Moreover, further applications of integrity tests, such as personnel development, profit from the findings about predictive and incremental predictive validity of integrity tests.

In this dissertation, integrity tests are shown to empirically predict OCB. This finding demonstrates the power of integrity tests to predict a variety of work-related behaviors. Several studies and meta-analysis had already revealed that integrity tests are highly accurate in predicting CWB and job performance. Now it has been shown that integrity tests additionally predict OCB, which may be important for personnel selection within social occupations such as nurses, priests, or kindergarten teachers. Furthermore, integrity tests can help identifying those employees that are involved with the company and its members. As a consequence, integrity tests can ensure that future employees support their company and help colleagues. Alongside personnel selection settings, the field of personnel development within companies could be another application for these tests: integrity tests could detect a need for OCB training in the company or verify OCB training by conducting before and post-training testing.

In addition, integrity tests were shown to predict OCB incrementally beyond two important personality correlates of conscientiousness and honesty-humility. This incremental predictive power of integrity tests was shown at the factor- and facet level of both personality traits. Despite a strong relationship between conscientiousness and OCB (Berry, Ones, & Sackett, 2007; Ilies et al., 2009), integrity tests had an additional gain in validity for predicting OCB. This finding is also important for personnel selection settings, in which only slightly higher ratings for OCB mean substantial gains for an organization and its members. Therefore, the valuable impact of incremental validity on the success of personnel selection and, as a result, on the success of a company is clear (Lievens et al., 2020).

Integrity tests were also shown to predict job performance incrementally over the two important personality correlates of conscientiousness and honesty-humility. Again, this incremental predictive power of integrity tests was shown at the factor- and facet levels of both personality traits. Conscientiousness must be considered one of the most important correlates of job performance (Barrick & Mount, 1991); only cognitive ability has comparable power to predict job performance as accurately (Schmidt & Hunter, 1998).

The finding that the application of integrity tests has incremental gains for predicting job performance is important for practice because integrity tests were found to have

additional benefits above other measures of personality for predicting job performance. Thus, predicting job performance as accurately as possible encourages companies to have productive employees and additionally reduces the additional cost of engaging new employees to replace those with a poor work ethic. Moreover, even small gains in validity for predicting and selecting future employees with beneficial job performance can make a significant difference not only for large companies with many employees and high sales but also for small companies with fewer employees and lower sales (Lievens et al., 2020).

7.3 STRENGTHS AND LIMITATIONS

This dissertation has provided new insight into the context of the construct of integrity as well as the criterion validity of integrity tests. Integrity tests have been reviewed with a focus on their construction and application. The main factors of the nomological network of integrity have been investigated in detail at their facet level to reveal new insights into the construct of integrity and the incremental validity of integrity tests has been widened. Finally, two promising rating methods have been tested to reduce faking on integrity tests.

All the studies reviewed in this doctoral thesis have contributed to five existing gaps in the research on integrity. First, after more than 20 years, the number of currently existing integrity tests has been identified. This up-to-date overview of integrity tests supports the orientation and choice of integrity tests. Second, a prototype integrity test was identified, helping to elucidate the confusing variety of integrity tests and improving the ability to generalize previous and new findings on integrity and integrity tests. Third, the factor- and facet level of honest-humility and OCB has been investigated in detail to advance the definition of integrity. Moreover, OCB was shown to be an important empirical correlate of integrity's nomological network. Fourth, integrity tests were shown to incrementally predict job performance as well as OCB over conscientiousness and honesty-humility. With regard to the application of integrity tests, these findings strengthen the predictive power of these kinds of tests in personnel selection settings. Fifth, the double-rating method has shown its potential to prevent faking on integrity tests. Moreover, the novel method is easy to apply and prevents faking before it occurs.

With regard to methodological aspects, this dissertation has a number of strengths: The use of both qualitative and quantitative research together with a correlative and experimental design offers a multi-methodological approach for this dissertation. Using multiple approaches has the advantage of counteracting methodological bias, which can distort data or their results. Moreover, the range of approaches and research designs used attempted to fill existing research gaps regarding integrity and integrity tests; for example, research on both novel methods to reduce faking was conducted with an experimental design as it was the first attempt to use these methods for integrity tests.

However, this dissertation has some limitations regarding methodological aspects: The use of only self-reported integrity measures constitutes a same-source bias: Using the same type of source, such as a self-reported test, can distort data and falsify results. For example, the problem of faking exists within self-reported integrity tests and contributes to the value of integrity. As a result, researchers emphasize the use of diverse sources, such as integrity ratings from supervisors or colleagues in addition to self-reported integrity measures. However, observer ratings of personality are also prone to faking (König, Steiner Thommen, Wittwer, & Kleinmann, 2017). In addition, self-reported integrity tests are the standard tool of integrity test research and are widely used, possibly because it is complex and difficult to collect data from other people, or to collect indirect data such as the theft rate in a company. Moreover, researchers found that self- and other-ratings of CWB are relatively comparable because the ratings generally result in very similar patterns of findings (Berry, Carpenter, & Barratt, 2012).

In addition to the same source bias caused by only using self-reported integrity tests, bias might be introduced by using only web-based data collections. For example, the level of faking within web-based surveys may be lower than in personal testing situations. As a result, the findings of reducing faking may be poorer due to the lower extent of faking. Nevertheless, previous studies revealed no difference between face-to-face and web-based data collections in terms of the incidence of faking (Gnambs & Kaspar, 2017; Rossiter, 2009). Moreover, tests are increasingly conducted in online, web-based formats due to lower costs and more flexible application (Woods et al., 2020). In line with this trend, the 10-item integrity test used in Chapters 4 and 5, was designed for use in web-based surveys and has a norm sample, which was collected online.

A further limitation could be found in the fact that the integrity tests used in this doctoral thesis were not established measures of integrity (i.e., the 10-item integrity test by Göritz, 2014, or the integrity-related items created for the study in Chapter 6). Nevertheless, the 10-item integrity test by Göritz (2014) used in Chapters 4 and 5 was carefully chosen with the aim of using a prototype integrity test that includes, for example, one dimension and a solid norm sample size. Given the wide diversity in integrity tests, the results of such a prototype integrity test can be better generalized because this type of test better represents the main points of the majority of integrity tests. The integrity-related items of Chapter 6 were created for this study to establish a solid basis from which to investigating faking. Many CWB actions may be specific to certain work contexts or job positions (e.g., embezzling money when working in the finance department of a company), so that some respondents may never have had the opportunity to engage in such CWBs (Bowling & Gruys, 2009). Therefore, within the integrity-related items, a decision about a minor delinquent action was integrated to generate higher rates and a full range of faked answers given.

The chapter on faking may be criticized for having no validation data for both rating methods to reduce faking. However, validation data could not be collected for three reasons: First, previous studies also revealed no validation data and, therefore, there was no opportunity to refer to previous results. Second, collecting validation data is an extensive process, which cannot be implemented for one part of one dissertation because the scope of this work is not broad enough to generate and present validation data. Third, the double-rating method is a novel, widely unknown method, originating from social psychology and adopted in integrity tests for the first time.

To strengthen the potential effect of rating methods to reduce faking, two different approaches were used for the investigation of the double-rating method: (a) the mean difference between method and control group, and (b) the correlation of the method with an impression management scale. These two approaches are widely used within faking research (e.g., Yu, 2008; McFarland & Ryan, 2000) and clearly indicate an effect of reduced faking.

The quality of the participants is an additional element of the discussion on the strengths and limitations of this dissertation. With regard to the data in Chapters 4 and 5, the participants are all part-time workers, which may cause a difference in the results of integrity testing. In previous studies, little difference was found between full-time workers

and part-time workers (e.g., Thorsteinson, 2003), but no study has investigated the difference between part- and full-time workers with reference to integrity tests. The relevant criterion seems not to be the time participants spend in their work environment; their length of service and the complexity of their work could be more important. A previous meta-analysis revealed a moderating effect of job complexity on integrity; that is, integrity is best predicted when job complexity is low or high (Ones et al. 1993). Moreover, the work experience of employees was shown to be positively related to integrity (Marcus, 2006): employees with longer experience tended to have greater integrity and engage in less CWB. This aspect of participants, their working experience, was included in all data collections in this dissertation.

With regard to Chapter 6, the data refer to 1,450 working participants. This population offers clear advantages for a study investigating integrity tests: First, such a large number of participants is rarely found in research studies. Second, integrity and integrity tests could be investigated in the primary target setting of the working environment. Despite their large working population, many research studies use a manageable number of students for reasons of simplicity. The working sample of this dissertation allows the findings to be generalized for the primary field of application of integrity tests. Moreover, the variation in demographical data within the populations of participants investigated may help to avoid undesirable effects, for example, that integrity tests correlate with the length of experience of employees (Marcus, 2006).

Regarding cultural elements, this dissertation referenced only German samples. This choice may prompt criticism and could limit the cultural generalizability of the findings because the results may differ from those found within other cultural samples. The available evidence for this assumption is mixed. Whereas some previous studies could not detect significant differences in integrity test scores for people from different cultural backgrounds (e.g., Fortmann, Leslie, & Cunningham, 2002; Marcus, Lee, & Ashton, 2007), other studies found differences in integrity test scores between different cultures (e.g., Fine, 2010; Ones, Wiernik, Viswesvaran, & Schmidt, 2014). Billings and Dages (2018) found that cultural differences in integrity test scores are related to the Corruption Perceptions Index scores for these countries.

Another element that is central to the discussion of cultural differences is the characteristics of the construction of integrity tests. While many integrity tests are available

in different languages, norm samples of people from different cultures revealed no differences in integrity test scores between these people (e.g., the Employee Reliability Inventory). It is important to note that the majority of integrity tests and the research on integrity and its tests stem from North America. Nevertheless, researchers found that integrity tests are valid outside of the United States for both CWBs and job performance (Billings & Dages, 2018). Moreover, findings can be replicated for German samples. For example, Marcus (2006) found that participant age has a positive correlation with integrity in his mixed (i.e., overt and personality-based measure) integrity test, IBES. This finding is in line with Ones and Viswesvaran (1998a) who found a positive correlation between age and integrity for three different overt integrity tests (i.e., Reid Report, Stanton, Personnel Selection Inventory).

7.4 DIRECTIONS FOR FUTURE RESEARCH

The empirical findings presented in this dissertation have filled some research gaps regarding the construct and criterion validity of integrity tests. Moreover, the current data illuminated important facets of personality and behavior related to integrity, thereby supporting the discussion of unresolved issues of the construct and validity of integrity. Based on these findings, the current research identifies the following areas for further research.

First, only overt integrity tests were used in this dissertation. Integrity tests can be categorized into two types, based on their item content (Sackett et al., 1989): Overt tests, which are obviously related to integrity, refer to the attitudes and admissions of CWB. Personality-based tests, which relate to personality factors underlying the construct of integrity, refer to personality factors (e.g., conscientiousness) and facets (e.g., sensation-seeking, a facet of extraversion). Both types of integrity test correlate with $\rho = .39$ (Ones et al., 1993), indicating a significant variation between the two types of test. In a meta-analysis, overt integrity tests were revealed to predict CWB more accurately than personality-based integrity tests (Ones et al., 1993).

All empirical studies in this dissertation only used overt integrity tests, while using a personality-based integrity test could expand the validity of the results found in these

studies. For some results, relationships may be diminished: For example, the relationship between OCB and integrity could be weakened by using a personality-based integrity test because, in line with previous findings regarding the prediction of CWB (Ones et al., 1993), behavior might have a lower correlation with a personality-based test than an overt integrity test. For other results, relationships may be strengthened: For example, the relationship between honesty-humility and integrity may be increased because personality traits should have a higher correlation with personality-based rather than overt integrity tests. Moreover, some results may differ from the current findings which used an overt integrity test: for example, other facets of honesty-humility or OCB may be more or less important when assessed through a personality-based test instead of an overt integrity test. Nevertheless, confirmation of the current results with both overt and personality-based integrity tests could strengthen these findings and shed new light on the construct of integrity and integrity tests.

Second, other traits could be used to provide insights into the construct of integrity. In line with the current research, two central avenues of research could be pursued: 1) Other personality traits could be investigated on the facet level to obtain a detailed view of the facet level of personality constructs surrounding integrity. 2) Personality traits that are also strongly relevant to integrity could be controlled for, and these findings could support the strength of incremental validity that integrity tests have. Aside from conscientiousness and honesty-humility used in the current research, emotional stability (Marcus et al., 1997; Ones, 1993) or agreeableness (J. Hogan & Brinkmeyer, 1997; Marcus et al., 1997; Ones, 1993), which also consist of a factor and facets, might be used. Self-control might be an especially promising trait to include in the design of future studies because it showed to correlate substantially with integrity (Sackett & Wanek, 1996) and even to add variance when predicting integrity over the Big Five traits (Bazzy et al., 2017). Moreover, controlling for the trait of self-control was found to diminish the relationship between conscientiousness and integrity (Bazzy et al., 2017).

Furthermore, the dark triad of personality (Paulhus, & Williams, 2002)–consisting of the personality traits of Machiavellianism, subclinical narcissism, and subclinical psychopathy–might provide new perspectives on the concept of integrity. These traits revealed relevant relationships with negative work-related behavior of employees, especially with CWB (Sackett et al., 2017). Moreover, a recent meta-analysis found that

honesty-humility, which is a highly important correlate of integrity, relates very strongly to the dark triad (Howard & Van Zandt, 2020). Interestingly and in contrast, integrity was not found to significantly correlate with psychopathy (Connelly et al., 2006). This finding might has been influenced by a manipulative interpersonal style, which is a key characteristic of all three traits (Paulhus & Williams, 2002). People with a high level of traits from the dark triad aim to change the behavior or perception of others through deceptive tactics for their own advantage. Appearing to have integrity might be positive for people with a high degree of dark triad traits. As a result, these people might be experts in faking, and a more indirect method of detecting faking, such as the double rating method, could be very helpful in this context.

Moreover, behavioral components must be integrated into the concept of integrity. The component of compliance, which was found within some integrity tests and within concepts of OCB, might be a key component of the construct of integrity. Identification of all the key components of integrity and the key core of integrity tests may help to achieve a clear understanding and definition of integrity and, further, to form integrity tests with a clearer and narrower focus.

Third, using other concepts of OCB may bring new insights into the relationship between OCB and integrity and into the stability of results. For this dissertation, the OCB taxonomy of Organ (1988) was used, including the facets of altruism, conscientiousness, sportsmanship, and civic virtue.[10] However, future research could also use the concept of Williams and Anderson (1991) or that of Van Scotter and Motowidlo (1996). Both groups of authors suggested two OCB dimensions: First, OCB with reference to other people, known as behavior directed toward individuals (OCBI) or interpersonal facilitation and, second, OCB with reference to organizational aspects, known as behavior directed toward the organization (OCBO) or job dedication. Interestingly, the facets of Organ's OCB concept can be categorized into both dimensions. The results may reveal interesting relationships between integrity and OCB towards individuals, in contrast to relationships between integrity and OCB towards organizations. There may be an advantage in conducting research in special application selection settings, where the prediction of good team work with other

[10] The fifth facet of courtesy was not integrated as a subscale in the questionnaire because the facet could not be replicated as a factor standing alone.

colleagues is very important (e.g., surgical teams). Moreover, it must be considered that this two-dimensional distinction of OCB also exists within a concept of CWB developed by Robinson and Bennett (1995). As a consequence, using this taxonomy of OCB fits with the primary prediction criterion of integrity tests and further supports the strength of the predictive and incremental validity of integrity tests.

Forth, it is urgently necessary to more strongly include new digital developments in the research on recruitment and selection (Woods et al., 2020) as well as on integrity tests. For example, no research has yet performed on integrity tests conducted on mobile phones or tablets. It therefore remains unclear, for example, whether a test application via mobile phone would work, what environmental variables would influence this application, and what the consequences would be for applicants and personnel selection procedures. To provide another example of a technical development, the approach of gamified assessments might bring a fresh perspective to the research on integrity tests, particularly because self-reported integrity tests and faking are still significant issues in research. There might be a handful of studies that use extensive scenarios programmed on computers in which the participant has to act like an employee and make decisions on integrity-relevant situations. Moreover, new technical developments and approaches exists in this field that are only just beginning to be used in research. Gamified integrity assessments within personnel selection procedures or organizational trainings could be used. Participants could explore virtual worlds by using virtual-reality glasses and interact with avatars who represent employees. These realistic work-related settings might reduce the problem of faking (Woods et al., 2020) and–as a behavioral measure–the problem of same source bias as a consequence of self-reported data.

Given that integrity tests are designed with the purpose of detecting CWB tendencies in future employees, the findings should be tested in the context of personnel selection. It may be interesting to discover whether the findings could be replicated with applicants in the setting of a personnel selection procedure. The question is whether the incremental prediction of job performance and OCB beyond conscientiousness and honesty-humility is still found to be significant at the same level. Moreover, the current results will bring strengths to the selection of future employees. For example, applicants can be selected who additionally show a tendence to OCB. To give another example, integrity tests with a new

rating format may have the potential to reduce faking; replicating the results within a selection setting offers the opportunity to expand the external validity of the findings.

Finally, it is important to mention that all data in this dissertation were collected before the coronavirus pandemic began. Results might thus change if data were collected now because new workplace requirements have been established, such as working from home. The topic of integrity testing as well as its construct and criterion validity would thus be seen in a new light and be even more relevant for the businesses of today.

7.5 CONCLUSION

This dissertation focused on two central questions: *What is integrity* and *how can integrity be used?* Guided by these fundamental questions, this doctoral thesis provided new insights into the construct of integrity and contributed to current research on the validity of integrity tests by identifying and addressing research gaps. These contributions were made via a variety of approaches, including qualitative and quantitative analyses as well as experimental and correlative designs.

The overview of integrity tests revealed the existence of 76 integrity tests. In exploring the question of what constitutes integrity, these integrity tests were shown to be heterogenous and to adopt many different constructs. Moreover, the benefits and drawbacks of the construction of integrity tests as well as their application criteria were identified and summarized in a prototype integrity test. This prototype has the potential to consolidate the confusing variety of content found in integrity tests; moreover it may help in forming a clear framework for research purposes and could support the generalization of former and new findings on integrity and integrity tests.

Furthermore, the findings indicate the importance of personality factors, which are reflected in the trait of honesty-humility, and of organizational behavior in the form of OCB. In addition, some facets of both constructs were found to be most important for integrity: modesty, fairness, sportsmanship, conscientiousness, and altruism. In terms of these findings on the key components of integrity, the construct validity of integrity tests is improved. All of these findings have implications relating to the question of what integrity is

and provide an impetus to redefine integrity by using both key personality traits and key behavioral constructs found at the factor- and facet levels.

To answer the question of how integrity is used, findings indicate that integrity tests have incremental validity for predicting both job performance and OCB beyond conscientiousness and honesty-humility at the factor- and facet level. These findings contribute to broadening the criterion validity of integrity tests.

Moreover, the double-rating method was shown to provide a promising approach to reducing faking on integrity tests and thereby to strengthening the criterion validity of these tests by diminishing error variance. The implications of these findings can be used in integrity tests with the double-rating method in different fields of business whenever job performance and OCB are relevant (e.g., in personnel development or in compliance- and awareness training).

Van Iddekinge et al.'s (2012a) claim that "… much more research is needed to increase understanding about what integrity tests measure and whether and how the underlying facets relate to valued criteria" (p. 520) was found to hold water. Moreover, this dissertation made an important contribution in response by reviewing integrity tests, expanding the nomological network of integrity, and reducing faking on integrity tests. In the light of these findings, this dissertation contributed to new knowledge about and a greater understanding of what integrity is and how it can be used.

REFERENCES

Academy of Human Resource Development (AHRD) (1999). *Standards on ethics and integrity.* Baton Rouge, LA: Academy of Human Resource Development. Retrieved from https://www.ahrd.org/page/standards [01.03.2021]

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl, & J. Beckmann (Eds.), *Action-control: From cognition to behavior* (pp. 11–39). Heidelberg: Springer.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. doi:10.1016/0749-5978(91)90020-T

Ajzen, I. (2017). Theory of planned behavior with background factors. *Retrieved from https://people.umass.edu/aizen/tpb.background.html* [01.03.2021]

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918. doi:10.1037/0033-2909.84.5.888

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, *22*, 453–474. doi:10.1016/0022-1031(86)90045-4

Allen, T. D., Barnard, S., Rush, M. C., & Russell, J. E. (2000). Ratings of organizational citizenship behavior: Does the source make a difference?. *Human Resource Management Review*, *10*(1), 97–114. doi:10.1016/S1053-4822(99)00041-8

Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, *60*(1), 59–72. doi:10.1177/00131640021970367

Ambrose, M. L., Seabright, M. A., & Schminke, M. (2002). Sabotage in the workplace: The role of organizational injustice. *Organizational Behavior and Human Decision Processes*, *89*(1), 947–965. doi:10.1016/S0749-5978(02)00037-7

Anglim, J., Lievens, F., Everton, L., Grant, S. L., & Marty, A. (2018). HEXACO personality predicts counterproductive work behavior and organizational citizenship behavior in low-stakes and job applicant contexts. *Journal of Research in Personality*, *77*, 11–20. doi:10.1016/j.jrp.2018.09.003

Ash, P. (1991). Comparison of two integrity tests based upon youthful or adult attitudes and experiences. *Journal of Business and Psychology*, *5*(3), 367–381.

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, *15*(5), 327–353. doi:10.1002/per.417

Ashton, M. C., & Lee, K. (2005). Honesty–Humility, the Big Five, and the Five-Factor Model. *Journal of Personality*, *73*, 1321–1353. doi:10.1111/j.1467-6494.2005.00351.x

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. doi:10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008). The prediction of honesty–humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality*, *42*(5), 1216–1228. doi:10.1016/j.jrp.2008.03.006

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. doi:10.1080/00223890902935878

Ashton, M. C., & Lee, K. (2010). Trait and source factors in HEXACO-PI-R self-and observer reports. *European Journal of Personality*, *24*(3), 278–289. doi:10.1002/per.759

Ashton, M. C., & Lee, K. (2016). Age trends in HEXACO-PI-R self-reports. *Journal of Research in Personality*, *64*, 102–111. doi:10.1016/j.jrp.2016.08.008

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., ... De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356. doi:10.1037/0022-3514.86.2.356

Association of Test Publishers (ATP) (2010). *Model guidelines for pre-employment integrity testing programs*, 3rd ed. Washington, DC: Author. Retrieved from: https://www.testpublishers.org/assets/documents/Model_Guidelines_Third_Edition.pdf [01.03.2021]

Audi, R., & Murphy, P. E. (2006). The many faces of integrity. *Business Ethics Quarterly*, *16*(1) 3–21. Retrieved from https://www.jstor.org/stable/3857724?casa_token=VML2Rw7rgZoAAAAA%3AKdweFC79D-2bSvvYHm3R4CWH3ZQ0VL96-AU2kPOWPJ4sF5Ds91jdR8mKqv5NwQGO84KwUUdEpllucRw_vK4-QP1GT7pnr5JzdhyhK8kEJK8KkY_O34&seq=1#metadata_info_tab_contents [01.03.2021]

Barnard, A., Schurink, W., & De Beer, M. (2008). A conceptual framework of integrity. *SA Journal of Industrial Psychology*, *34*(2), 40–49. doi: 10.1177/1534484313492333

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: Aa meta-analysis. *Personnel Psychology*, *44*(1), 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x

Bazzy, J. D., Woehr, D. J., & Borns, J. (2017). An examination of the role of self-control and impact of ego depletion on integrity testing. *Basic and Applied Social Psychology*, *39*(2), 101–111. doi.org/10.1080/01973533.2017.1283502

Bechtold, M. N., Welk, C., Zapf, D., & Hartig, J. (2007). Main and moderating effects of self-control, organizational justice, and emotional labour on counterproductive behaviour at work. *European Journal of Work and Organizational Psychology*, *16*(4), 479–500. doi:10.1080/13594320701662618

Beck, L., & Ajzen, I. (1991). Predicting dishonest actions using the theory of planned behavior. *Journal of Research in Personality*, *25*(3), 285–301. doi:10.1016/0092-6566(91)90021-H

Becker, T. E. (1998). Integrity in organizations: Beyond honesty and conscientiousness. *Academy of Management Review*, *23*(1), 154–161. doi:10.5465/amr.1998.192969

Becker, T. E. (2005). Development and validation of a situational judgment test of employee integrity. *International Journal of Selection and Assessment*, *13*(3), 225–232. doi.org/10.1111/j.1468-2389.2005.00319.x

Becker, G. K. (2009). Integrity as moral ideal and business benchmark. *Journal of International Business Ethics*, *2*(2), 70–84. Retrieved from http://cibe.org.cn/uploadfile/otherfile/jibe-09-v2-n2.pdf#page=70 [01.03.2021]

Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, *85*(3), 349–360. doi:10.1037/0021-9010.85.3.349

Bennett, R. J., & Stamper, C. L. (2001). *Corporate citizenship and deviancy: A study of discretionary work behavior.* Manuscript submitted for publication. doi.org/10.1016/S1074-7877(02)03015-5

Bergmann, T. J., Mundt, D. H., & Illgen, E. J. (1990). The evolution of honesty tests and means for their evaluation. *Employee Responsibilities and Rights Journal*, *3*(3), 215–223. doi.org/10.1007/BF01396574

Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, *97*(3), 613–636. doi:10.1037/a0026739

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*(2), 410–424. doi:10.1037/0021-9010.92.2.410

Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology*, *62*(4), 833–863. doi:10.1111/j.1744-6570.2009.01159.x

Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, *60*(2), 271–301. doi:10.1111/j.1744-6570.2007.00074.x

Betts, G. L. (1947). Biographical case history. *Minneapolis, MN: Educational Test Bureau*.

Billings, S. W., & Dages, K. D. (2018). Cross-cultural validity of integrity assessments for lower-level and higher-level jobs. *International Journal of Selection and Assessment*, *26*(1), 66–74. doi:10.1111/ijsa.12198

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317–335. doi:10.1111/j.1468-2389.2006.00354.x

Bock, M. (2007). *Kriminologie* [*Criminology*]. Mainz: Vahlen. Retrieved from https://books.google.de/books?hl=de&lr=&id=qzxiDwAAQBAJ&oi=fnd&pg=PR3&dq=Bock,+M.+(2007).+Kriminologie++Mainz:+Vahlen..+&ots=MAiXSTJYKS&sig=kN804v-CHjSVO-T4eGe-Dq5oiXE#v=onepage&q&f=false [01.03.2021]

Borman, W. C., & Motowidlo, S. M. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel Selection in Organizations (*pp. 71–98). San Francisco: Jossey-Bass.

Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, *9*(1-2), 52–69. doi:10.1111/1468-2389.00163

Bourdage, J. S., Lee, K., Lee, J. H., & Shin, K. H. (2012). Motives for organizational citizenship behavior: Personality correlates and coworker ratings of OCB. *Human Performance*, *25*(3), 179–200. doi:10.1080/08959285.2012.683904

Bowling, N. A. (2010). Effects of job satisfaction and conscientiousness on extra-role behaviors. *Journal of Business and Psychology*, *25*(1), 119–130. doi:10.1007/s10869-009-9134-0

Bowling, N. A., & Gruys, M. L. (2010). Overlooked issues in the conceptualization and measurement of counterproductive work behavior. *Human Resource Management Review*, *20*(1), 54–61. doi:10.1016/j.hrmr.2009.03.008

Brooks, P. B., & Arnold, D. A. (1989). *Reid Report examiner's manual, (3rd ed.)*. Chicago: Reid Psychological Systems.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance*, *24*(4), 358–372. doi:10.1080/08959285.2011.597473

Burns, G. N., Shoda, E. A., & Roebke, M. A. (2017). Putting applicant faking effects on personality tests into context. *Journal of Managerial Psychology*. *32*(6), 460–468. doi:10.1108/JMP-01-2017-0031

Camara, W. J., & Schneider, D. L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist*, *49*(2), 112–119. doi:10.1037/0003-066X.49.2.112

Casillas, A., Robbins, S., McKinniss, T., Postlethwaite, B., & Oh, I. S. (2009). Using narrow facets of an integrity test to predict safety: A test validation study. *International Journal of Selection and Assessment*, *17*(1), 119–125. doi:10.1111/j.1468-2389.2009.00456.x

Cassel, R. N., & Betts, G. L. (1956). The development and validation of a life experience inventory for the identification of delinquency of prone youth. In *American Psychologist, 11*(8), 366–366.

Catano, V. M., O'Keefe, D. F., Francis, R. E., & Owens, S. M. (2018). Construct-based approach to developing a short, personality-based measure of integrity. *International Journal of Selection and Assessment*, *26*(1), 75–92. doi:10.1111/ijsa.12197

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, *37*(4), 319–338. doi:10.1016/S0092-6566(02)00578-0

Chiaburu, D. S., & Lim, A. S. (2008). Manager trustworthiness or interactional justice? Predicting organizational citizenship behaviors. *Journal of Business Ethics*, *83*(3), 453–467. doi:10.1007/s10551-007-9631-x

Chiaburu, D. S., Oh, I.-S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, 96(6), 1140–1166. doi:10.1037/a0024004

Christiansen, N. D., & Robie, C. (2011). Further consideration of the use of narrow trait scales. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *43*(3), 183–194. doi:10.1037/a0023069

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *86*(2), 278–321. doi:10.1006/obhd.2001.2958

Connelly, B. S., Lilienfeld, S. O., & Schmeelk, K. M. (2006). Integrity tests and morality: Associations with ego development, moral reasoning, and psychopathic personality. *International Journal of Selection and Assessment*, *14*(1), 82–86. doi:10.1111/j.1468-2389.2006.00335.x

Costa, P. T., & 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.

Coyne, I., & Bartram, D. (2002). Assessing the effectiveness of integrity tests: A review. *International Journal of Testing*, *2*(1), 15–34. doi:10.1207/S15327574IJT0201_2

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological Tests and Personnel Decisions*, 2nd, University of Illinois Press, Urbana, IL.

Crowne. D. P., & Marlowe. D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology. 24,* 349–354. doi:10.1037/h0047358

Crutzen, R., & Göritz, A. S. (2011). Does social desirability compromise self-reports of physical activity in web-based research?. *International Journal of Behavioral Nutrition and Physical Activity*, *8*(1), 1–4. doi:10.1186/1479-5868-8-31

Cullen, M. J., & Sackett, P. R. (2003). *Integrity testing in the workplace: I/O Assessment*, 149–165. In J. C. Thomas (Ed.), *Comprehensive Handbook of Psychological Assessment. Volume IV: I/O Assessment* (Vol. 4). New York: John Wiley and Sons.

Cunningham, M. R., & Ash, P. (1988). The structure of honesty: Factor analysis of the Reid Report. *Journal of Business and Psychology*, *3*(1), 54–66. doi:10.1007/BF01016748

Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*, *90*(6), 1241–1255. doi:10.1037/0021-9010.90.6.1241

Dalal, D. K., & Hakel, M. D. (2016). Experimental comparisons of methods for reducing deliberate distortions to self-report measures of sensitive constructs. *Organizational Research Methods*, *19*(3), 475–505. doi:10.1177/1094428116639131

De Clercq, D., Haq, I. U., & Azeem, M. U. (2019). Time-related work stress and counterproductive work behavior. *Personnel Review*, *48*(7), 1756–1781. doi:10.1108/PR-07-2018-0241

De Meijer, L. A., Born, M. P., Van Zielst, J., & Van der Molen, H. T. (2010). Construct-driven development of a video-based situational judgment test for integrity. *European Psychologist, 15*(3), 229–236. doi:10.1027/1016-9040/a000027

De Vries, A., De Vries, R. E., & Born, M. P. (2011). Broad versus narrow traits: Conscientiousness and Honesty–Humility as predictors of academic criteria. *European Journal of Personality*, *25*(5), 336–348. doi:10.1002/per.795

De Vries, R. E., & Van Gelder, J. L. (2013). Tales of two self-control scales: Relations with five-factor and HEXACO traits. *Personality and Individual Differences*, *54*(6), 756–760. doi:10.1016/j.paid.2012.12.023

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, *91*(1), 40–57. doi:10.1037/0021-9010.91.1.40.

Dwight, S. A., & Alliger, G. M. (1997). Reactions to overt integrity test items. *Educational and Psychological Measurement*, *57*(6), 937–948. doi:10.1177/0013164497057006004

Eignor, D. R. (2013). The standard for educational and psychological testing. In *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 245-250). American Psychological Association.

Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology*, *92*(2), 386. doi.org/10.1037/0021-9010.92.2.386

Evers, A. (2001). The revised Dutch rating system for test quality. *International Journal of Testing*, *1*(2), 155–182. doi:10.1207/S15327574IJT0102_4

Feather, N. T., & Rauter, K. A. (2004). Organizational citizenship behaviours in relation to job status, job insecurity, organizational commitment and identification, job satisfaction and work values. *Journal of Occupational and Organizational Psychology*, *77*(1), 81–94. doi:10.1348/096317904322915928

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: SAGE Publications.

Fine, S., & Gottlieb-Litvin, Y. (2013). Justifying counterproductive work behaviors and an integrity-based conditional reasoning test: Back to the drawing board?. *International Journal of Selection and Assessment*, *21*(3), 328–333. doi:10.1111/ijsa.12042

Fine, S., & Pirak, M. (2016). Faking fast and slow: Within-person response time latencies for measuring faking in personnel testing. *Journal of Business and Psychology*, *31*(1), 51–64. doi:10.1007/s10869-015-9398-5

Fischer, D., & Bates, J. A. (2008, April). *The development of an IAT for workplace integrity*. Paper presented at the 23rd Annual Society for Industrial and Organizational Psychology, San Francisco, California.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the marlowe-crowne social desirability scale. *Educational and Psychological Measurement, 53*(2), 417–424. doi:10.1177/0013164493053002011

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, *20*(2), 303–315. doi:10.1086/209351

Forster, M. R. (2000). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology*, *44*(1), 205–231. doi:10.1006/jmps.1999.1284

Fox, S., Spector, P. E., Goh, A., & Bruursema, K. (2007). Does your coworker know what you're doing? Convergence of self-and peer-reports of counterproductive work behavior. *International Journal of Stress Management*, *14*(1), 41–60. doi:10.1037/1072-5245.14.1.41

Gerber-Braun, B. (2010). *The Double Cross: Individual differences between respondents with different response sets and styles on questionnaires* (Doctoral dissertation), München, Fakultät für Psychologie und Pädagogik der Ludwig–Maximilians–Universität). Retrieved from https://edoc.ub.uni-muenchen.de/11795/1/Gerber-Braun_Beatrice.pdf

Giordano, C., Ones, D. S., & Viswesvaran, C. (2020). Integrity testing and counterproductive work behavior. *The Wiley Encyclopedia of Personality and Individual Differences: Clinical, Applied, and Cross-Cultural Research*, 539-544. doi.org/10.1002/9781119547181.ch354

Gnambs, T., & Kaspar, K. (2017). Socially desirable responding in web-based questionnaires: A meta-analytic review of the candor hypothesis. *Assessment*, *24*(6), 746–762. doi:10.1177/1073191115624547

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology/Psychologie canadienne*, *50*(3), 151–160. doi:10.1037/a0015946

Göritz, A. S. (2014). A 10-item integrity test. Unpublished manuscript, Occupational and Consumer Psychology, University of Freiburg, Freiburg, Germany.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.

Gough, H. G. (1957). Manual for the California personality inventory. *Palo Alto, CA: Consulting Psychologists Press.*

Gough, H. G. (1971). *Preliminary manual for the "Personnel Reaction Blank."*. *Palo Alto, CA: Consulting*.

Graham, J. W. (1991). An essay on organizational citizenship behavior. *Employee Responsibilities and Rights Journal*, *4*(4), 249–270. doi:10.1007/BF01385031

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, *75*(5), 561–568. doi:10.1037/0021-9010.75.5.561

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, *36*(3), 341–355. doi:10.1108/00483480710731310

Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology*, *1*(3), 308–311. doi:10.1111/j.1754-9434.2008.00053.x

Griffith, R. L., & Peterson, M. H. (2011). One piece at a time: The puzzle of applicant faking and a call for theory. *Human Performance*, *24*(4), 291–301. doi:10.1080/08959285.2011.597474

Gruys, M. L., & Sackett, P. R. (2003). Investigating the dimensionality of counterproductive work behavior. *International Journal of Selection and Assessment*, *11*(1), 30–42. doi:10.1111/1468-2389.00224

Harris, W. G. (1987). A component analysis of a pre-employment integrity measure. *Psychological Reports, 60*(3), 1051–1055. doi:10.2466/pr0.1987.60.3c.1051

Harris, W. G., Jones, J. W., Klion, R., Arnold, D. W., Camara, W., & Cunningham, M. R. (2012). Test publishers' perspective on "An updated meta-analysis": Comment on Van Iddekinge, Roth, Raymark, and Odle-Dusseau (2012). doi:10.1037/a0024767

Harris, M. M., & Sackett, P. R. (1987). A factor analysis and item response theory analysis of an employee honesty test. *Journal of Business and Psychology*, *2*(2), 122–135. doi:10.1007/BF01014207

Hastings, S. E., & O'Neill, T. A. (2009). Predicting workplace deviance using broad versus narrow personality variables. *Personality and Individual Differences*, *47*(4), 289–293. doi:10.1016/j.paid.2009.03.015

Hathaway, S. R. & McKinley, J. C. (1951). *The Minnesota Multiphasic Personality Inventory manual* (Rev. Ed.). New York: Psychological Corporation.

Heneman, R. C. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology*, 39, 811–826. doi:10.1111/j.1744-6570.1986.tb00596.x

Hertel, G., Bretz, E., & Moser, K. (2000). Freiwilliges Arbeitsengagement: Begriffsklärung und Forschungsstand. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, *31*(2), 121–140. doi:10.1007/s11612-000-0013-4

Hirschi, T., & Gottfredson, M. R. (2000). In defense of self-control. *Theoretical Criminology*, *4*(1), 55–69. doi:10.1177/1362480600004001003

Hogan, R. (1986). *Manual for the Hogan Personality Inventory*. Minneapolis: National Computer Systems.

Hogan, J., & Brinkmeyer, K. (1997). Bridging the gap between overt and personality-based integrity tests. *Personnel Psychology*, *50*(3), 587–599. doi:10.1111/j.1744-6570.1997.tb00704.x

Hogan, J., & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied psychology*, *74*(2), 273–279. doi:10.1037/0021-9010.74.2.273

Hogan, J., & Ones, D. S. (1997). Conscientiousness and integrity at work. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 849–870). New York, NY: Academic Press. doi.org/10.1016/B978-012134645-4/50033-0

Hollinger, R. C., & Clark, J. P. (1982). Formal and informal social controls of employee deviance. *The Sociology Quarterly, 23*(3), 333–343. doi.org/10.1111/j.1533-8525.1982.tb01016.x

Hollinger, R. C., & Clark, J. P. (1983). Deterrence in the workplace: Perceived certainty, perceived severity and employee theft. *Social Forces, 62*, 398–418. doi:10.1093/sf/62.2.398

Honts, C. R. (1991). The emperor's new clothes: Application of polygraph tests in the American workplace. *Forensic Reports*, *4*(2), 91–116. Retrieved from https://www.researchgate.net/profile/Charles-Honts/publication/232517551_The_emperor%27s_new_clothes_Application_of_polygraph_tests_in_the_American_workplace/links/5451e5bb0cf2bf864cbaa48d/The-emperors-new-clothes-Application-of-polygraph-tests-in-the-American-workplace.pdf [01.03.2021]

Hoon, H., & Tan, T. M. L. (2008). Organizational citizenship behavior and social loafing: The role of personality, motives, and contextual factors. *The Journal of Psychology*, *142*(1), 89–108. doi:10.3200/JRLP.142.1.89-112

Hough, L. M., & Schneider, R. J. (1996). Personality traits, taxonomies, and applications in organizations. In Murphy, K. (Ed.), *Individual differences and behavior in organizations* (pp. 31–88). San Francisco: Jossey-Bass. Howard, M. C., & Van Zandt, E. C. (2020). The discriminant validity of honesty-humility: A meta-analysis of the HEXACO, Big Five, and Dark Triad. *Journal of Research in Personality*, *87*, 103982. doi:10.1016/j.jrp.2020.103982

Hui, C. H. (2001). Double Rating as a Method to Encourage Candid Responses to Self-Report Instruments 1. *Journal of Applied Social Psychology*, *31*(1), 21–30. doi:10.1111/j.1559-1816.2001.tb02479.x

Ilies, R., Fulmer, I. S., Spitzmuller, M., & Johnson, M. D. (2009). Personality and citizenship behavior: The mediating role of job satisfaction. *Journal of Applied Psychology*, *94*(4), 945–959. doi:10.1037/a0013329

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution?. *Human Performance*, *13*(4), 371–388. doi:10.1207/S15327043HUP1304_3

James, L. R., & McIntyre, M. D. (2000). *Conditional reasoning test of aggression: Test manual*. Innovative Assessment Technology.

James, L. R., McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M., ... Williams, L. J. (2005). A conditional reasoning measure for aggression. *Organizational Research Methods*, *8*(1), 69–99. doi:10.1177/1094428104272182

Jeong, Y. R., Christiansen, N. D., Robie, C., Kung, M. C., & Kinney, T. B. (2017). Comparing applicants and incumbents: Effects of response distortion on mean scores and validity of personality measures. *International Journal of Selection and Assessment*, *25*(3), 311–315. doi:10.1111/ijsa.12182

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*(1), 1–19. doi:10.1207/S15327906MBR3501_1

Johnson, M. K., Rowatt, W. C., & Petrini, L. (2011). A new trait on the market: Honesty–Humility as a unique predictor of job performance ratings. *Personality and Individual differences*, *50*(6), 857–862. doi:10.1016/j.paid.2011.01.011

Jonason, P. K., & McCain, J. (2012). Using the HEXACO model to test the validity of the Dirty Dozen measure of the Dark Triad. *Personality and Individual Differences*, *53*(7), 935–938. doi:10.1016/j.paid.2012.07.010

Jones, J. W., Brasher, E. E., & Huff, J. W. (2002). Innovations in integrity-based personnel selection: Building a technology-friendly assessment. *International Journal of Selection and Assessment*, *10*(1-2), 87–97. doi.org/10.1111/1468-2389.00195

Jones, J. W., & Terris, W. (1984). Validity and factor structure of a dishonesty test that measures attitudes toward theft. *Unpublished technical report. Park Ridge, IL: London House*.

Jones, J. W., & Terris, W. (1991). Integrity testing for personnel selection: An overview. *Forensic Reports*.

Kaptein, S. P. (2008). Developing a measure of unethical behavior in the workplace: a stakeholder perspective. *Journal of Management, 34*(5), 978–1008. doi:10.1177/0149206308318614

Karren, R. J., & Zacharias, L. (2007). Integrity tests: Critical issues. *Human Resource Management Review*, *17*(2), 221–234. doi:10.1016/j.hrmr.2007.03.007

Kasten, N., Freund, P. A., & Staufenbiel, T. (2020). "Sweet little lies": An in-depth analysis of faking behavior on Situational Judgment Tests compared to personality questionnaires. *European Journal of Psychological Assessment*, *36*(1), 136–148. doi:10.1027/1015-5759/a000479

Khalid, S. A., & Ali, H. (2005). Self and superior ratings of organizational citizenship behavior: Are there differences in the source of ratings?. *Problems and Perspectives in Management*, (4), 147–153. Retrieved from file:///C:/Users/NADINE~1/AppData/Local/Temp/prperman_2005_4_15.pdf [01.03.2021]

Kluemper, D. H., DeGroot, T., & Choi, S. (2013). Emotion management ability: Predicting task performance, citizenship, and deviance. *Journal of Management*, *39*(4), 878–905. doi:10.1177/0149206311407326

König, C. J., Steiner Thommen, L. A., Wittwer, A. M., & Kleinmann, M. (2017). Are observer ratings of applicants' personality also faked? Yes, but less than self-reports. *International Journal of Selection and Assessment*, *25*(2), 183–192. doi:10.1111/ijsa.12171

Laginess, A. J. (2016). *Mapping Integrity in the Domain of Trait Personality*. Available from FIU Electronic Theses and Dissertations. (3365) doi:10.25148/etd.FIDC001969

Lamnek, S. (1993). Qualitative Sozialforschung [Qualitative social research]. 2nd ed. Weinheim: Beltz.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. doi:10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and Narcissism in the Five-Factor Model and the HEXACO model of personality structure. *Personality and Individual Differences*, *38*(7), 1571–1582. doi:10.1016/j.paid.2004.09.016

Lee, K., & Ashton, M. C. (2013). Prediction of self-and observer report scores on HEXACO-60 and NEO-FFI scales. *Journal of Research in Personality*, *47*(5), 668–675. doi:10.1016/j.jrp.2013.06.002

Lee, K., Ashton, M. C., & De Vries, R. E. (2005). Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure. *Human Performance*, *18*(2), 179–197. doi:10.1207/s15327043hup1802_4

Lee, K., Ashton, M. C., Morrison, D. L., Cordery, J., & Dunlop, P. D. (2008). Predicting integrity with the HEXACO personality model: Use of self-and observer reports. *Journal of Occupational and Organizational Psychology*, *81*(1), 147–167. doi:10.1348/096317907X195175

Lee, K., Ashton, M. C., & Shin, K. H. (2005). Personality correlates of workplace anti-social behavior. *Applied Psychology*, *54*(1), 81–98. doi:10.1111/j.1464-0597.2005.00197.x

Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the Dark Triad and Honesty–Humility. *European Journal of Personality*, *27*(2), 169–184. doi:10.1002/per.1860

Lee, Y., Berry, C. M., & Gonzalez-Mulé, E. (2019). The importance of being humble: A meta-analysis and incremental validity analysis of the relationship between honesty-humility and job performance. *Journal of Applied Psychology*, *104*(12), 1535–1546. doi.org/10.1037/apl0000421

Lensvelt-Mulders, G. J. L. M., Hox, J. J., Van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods & Research*, *33*(3), 319-348. doi.org/10.1177/0049124104268664

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.

Lievens, F., Sackett, P. R., & Zhang, C. (2020). Personnel selection: a longstanding story of impact at the individual, firm, and societal level. *European Journal of Work and Organizational Psychology*, 1–12. doi:10.1080/1359432X.2020.1849386

Llobet, J. M. (2009). Applicant Risk Profiler. Administrator's Manual. HR Assessments. Retrieved from https://www.pdffiller.com/jsfiller-desk11/?requestHash=0a17983

bfd6fbe26adb020bf057df3fce489e4161f50faba269e3f18c201596b&projectId=688348495#7c8fe5ceee3988c04c0f5705f5b92284 [01.03.2021]

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

Marcus, B. (2006). *IBES – Manual zum Inventar berufsbezogener Einstellungen und Selbsteinschätzungen* [IBES – Manual of Inventory of Job-Related Attitudes and Self-Evaluation] (1st Ed.) Göttingen: Hogrefe.

Marcus, B., Ashton, M. C., & Lee, K. (2013). A note on the incremental validity of integrity tests beyond standard personality inventories for the criterion of counterproductive behaviour. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, *30*(1), 18–25. doi:10.1002/cjas.1235

Marcus, B., Bošnjak, M., Lindner, S., Pilischenko, S., & Schütz, A. (2007). Compensating for low topic interest and long surveys: a field experiment on nonresponse in web surveys. *Social Science Computer Review*, *25*(3), 372–383. doi:10.1177/0894439307297606

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*(1), 2–17.

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*(2), 113–130. doi:10.1111/j.1468-2389.2006.00338.x

Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big Five, or one in addition?. *Personnel Psychology*, *60*(1), 1–34. doi:10.1111/j.1744-6570.2007.00063.x

Marcus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology, 89*(4), 647–660. doi:10.1037/0021-9010.89.4.647

Marcus, B., Taylor, O. A., Hastings, S. E., Sturm, A., & Weigelt, O. (2016). The structure of counterproductive work behavior: A review, a structural meta-analysis, and a primary study. *Journal of Management*, *42*(1), 203–233. doi:10.1177/0149206313503019

Marcus, B., Te Nijenhuis, J., Cremers, M., & Heijden-Lek, K. V. D. (2016). Tests of integrity, HEXACO personality, and general mental ability, as predictors of integrity ratings in the Royal Dutch Military Police. *International Journal of Selection and Assessment*, *24*(1), 63–70. doi:10.1111/ijsa.12130

Mastrangelo, P. M., & Jolton, J. A. (2001). Predicting on-the-job substance abuse with a written integrity test. *Employee Responsibilities and Rights Journal*, *13*(2), 95–106. doi:10.1023/A:1014538309476

Matza, D. (1964). *Delinquency and Drift*. New York: John Wiley & Son.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. doi:10.5465/amr.1995.9508080335

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, *60*(1), 63–91. doi:10.1111/j.1744-6570.2007.00065.x

McFall, L. (1987). Integrity. *Ethics*, *98*(1), 5–20. doi:10.1086/292912

McFarland, C., & Miller, D. T. (1990). Judgments of self-other similarity–just like other people, only more so. *Personality and Social Psychology Bulletin*, *16*(3), 475–484. doi:10.1177/0146167290163006

McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment*, *11*(4), 265–276. doi:10.1111/j.0965-075X.2003.00250.x

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*(5), 812–821. doi:10.1037/0021-9010.85.5.812

McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking Behavior. *Journal of Applied Social Psychology*, *36*(4), 979–1016. doi:10.1111/j.0021-9029.2006.00052.x

McFarland, L. A., Ryan, A. M., & Kriska, S. D. (2003). Impression management use and effectiveness across assessment methods. *Journal of Management*, *29*(5), 641–661. doi:10.1016/S0149-2063_03_00030-8

Midot, Ltd. (2007). IntegriTEST test manual 2.0. Midot, Ltd.

Miller, R. L., Griffin, M. A., & Hart, P. M. (1999). Personality and organizational health: The role of conscientiousness. *Work & Stress*, *13*(1), 7–19. doi:10.1080/026783799296156

Motowidlo, S. J. (2003). *Job performance.* In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology, Vol. 12* (p. 39–53). John Wiley & Sons Inc. doi:10.1002/0471264385

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, *10*(2), 71–83. doi:10.1207/s15327043hup1002_1

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, *79*(4), 475–480. doi:10.1037/0021-9010.79.4.475

Mueller-Hanson, R., Heggestad, E. D., & Thornton III, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, *88*(2), 348–355. doi:10.1037/0021-9010.88.2.348

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests [Monograph.]. *Journal of Experimental Social Psychology, 21*(3), 262–283. doi:10.1016/0022-1031(85)90020-4

Mumford, M. D., Connelly, M. S., Helton, W. B., Strange, J. M., & Osburn, H. K. (2001). On the construct validity of integrity tests: Individual and situational factors as predictors of test performance. *International Journal of Selection and Assessment, 9*(3), 240–257. doi:10.1111/1468-2389.00177

Murphy, K. R. (2000). What constructs underlie measures of honesty or integrity?. In *Problems and solutions in human assessment* (pp. 265–283). Springer, Boston, MA. doi:10.1007/978-1-4615-4397-8_12

Murphy, K. R., & Lee, S. L. (1994a). Does conscientiousness explain the relationship between integrity and job performance?. *International Journal of Selection and Assessment*, *2*(4), 226–233. doi:10.1111/j.1468-2389.1994.tb00144.x

Murphy, K. R., & Lee, S. L. (1994b). Personality variables related to integrity test scores: The role of conscientiousness. *Journal of Business and Psychology*, *8*(4), 413–424. doi:10.1007/BF02230957

Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit [An inventory for the assessment of two factors of social desirability]. *Diagnostica, 48*, 121–129. doi:10.1026/0012-1924.48.3.121

Nerdinger, F. W. (2008). *Unternehmensschädigendes Verhalten erkennen und verhindern [Recognizing and preventing company damaging behavior]. Praxis der Personalpsychologie*. Göttingen: Hogrefe.

Nerdinger, F. W., Blickle, G., & Schaper, N. (2008). *Arbeits- und Organisationspsychologie [Work and organizational psychology]*. Heidelberg: Springer Medizin.

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, *13*(4), 250–260. doi:10.1111/j.1468-2389.2005.00322.x

Nicol, A. A. (1999). *A measure of workplace honesty*. Faculty of Graduate Studies, University of Western Ontario.

Nicol, A. A., & Paunonen, S. V. (2002). Validity evidence for the different item styles of overt honesty measures. *Journal of Business and Psychology*, *16*(3), 431–445. doi:10.1023/A:1012877008434

Noelliste, M. (2013). Integrity: An intrapersonal perspective. *Human Resource Development Review*, *12*(4), 474–499. doi:10.1177/1534484313492333

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: big five correlates of GPA and SAT scores. *Journal of Personality andSsocial Psychology*, *93*(1), 116. doi:10.1037/0022-3514.93.1.116

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

O'Brien, K. E., & Allen, T. D. (2007). The relative importance of correlates of organizational citizenship behavior and counterproductive work behavior using multiple sources of data. *Human Performance*, *21*(1), 62–88. doi:10.1080/08959280701522189

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual differences*, *43*(5), 971–990. doi:10.1016/j.paid.2007.03.017

O'Neill, T. A., Lewis, R. J., & Carswell, J. J. (2011). Employee personality, justice perceptions, and the prediction of workplace deviance. *Personality and Individual Differences*, *51*(5), 595–600. doi:10.1016/j.paid.2011.05.025

Ones, D. S. (1993). *Establishing construct validity for integrity tests* (Unpublished doctoral dissertation, University of Iowa, College of Business Administration). Retrieved from https://apps.dtic.mil/sti/pdfs/ADA293940.pdf [01.03.2021]

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, *17*(6), 609–626. doi:10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K

Ones, D. S., & Viswesvran, C. (1998a). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83*(1), 35–42. doi:10.1037/0021-9010.83.1.35

Ones, D. S., & Viswesvran, C. (1998b). Integrity testing in organizations. In R. W. Griffin, A. O'Leary, & J. M. Collins (Eds.), *Dysfunctional behavior in Organizations (*43-276): Vol. 2. Nonviolant Behaviors in Oranizations. London: JAI Press. Retrieved from https://www.researchgate.net/profile/Chockalingam-vish-Viswesvaran-2/publication/232460337_Integrity_testing_in_organizations/links/59e4d8f8458515250246e641/Integrity-testing-in-organizations.pdf [01.03.2021]

Ones, D. S., & Viswesvaran, C. (1998c). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, *11*(2-3), 245–269. doi:10.1080/08959285.1998.9668033

Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment, 9*(1-2), 31–39. doi:10.1111/*1468*-2389.00161

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*(6), 660–679. doi:10.1037/0021-9010.81.6.660

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, *78*(4), 679–703. doi:10.1037/0021-9010.78.4.679

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: a meta-analysis of integrity tests. *European Journal of Personality*, *17*(1), 19–38. doi:10.1002/per.487

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2012). Integrity tests predict counterproductive work behaviors and job performance well: Comment on Van Iddekinge, Roth, Raymark, and Odle-Dusseau (2012). doi:10.1037/a0024825

Ones, D. S., Viswesvaran, C., Schmidt, F. L., & Reiss, A. D. (1994, August). The validity of honesty and violence scales of integrity tests in predicting violence at work. In *annual meeting of the Academy of Management, Dallas, TX*.

Ones, D. S., Wiernik, B. M., Viswesvaran, C., & Schmidt, F. L. (2014, May). Cross-cultural validity of pre-employment integrity tests: A transcultural meta-analytic investigation. In D. S. Ones & B. M. Wiernik (Co-Chairs), Evidence-based practice in preemployment integrity testing: Validity, applicability, and usability. Symposium conducted at the annual conference of the Society for Industrial Organizational Psychology, Honolulu, HI.

Oostrom, J. K., Köbis, N. C., Ronay, R., & Cremers, M. (2017). False consensus in situational judgment tests: What would others do?. *Journal of Research in Personality*, *71*, 33–45. doi:10.1016/j.jrp.2017.09.001

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington Books/DC Heath and Com.

Organ, D. W. (1990). The motivational basis of organizational citizenship behavior. In B. M. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior* (pp. 43–72). Greenwich, CT: JAI Press.

Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, *10*(2), 85–97. doi:10.1207/s15327043hup1002_2

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, *48*(4), 775–802. doi:10.1111/j.1744-6570.1995.tb01781.x

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598–609. doi:10.1037/0022-3514.46.3.598

Paulhus, D. L. (1988). *Assessing self-deception and impression management in self-reports: The Balanced Inventory of Desirable Responding*. Unpublished manual, University of British Columbia, Vancouver Canada.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Paulhus, D. L. (1994). Balanced Inventory of Desirable Responding: Reference manual for BIDR version 6. Unpublished manuscript, University of British Columbia, Vancouver, Canada.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality. *Journal of Research in Personality, 36* (6), 556–563. doi:10.1016/S0092-6566(02)00505-6

Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences*, *37*(6), 1137–1151. doi:10.1016/j.paid.2003.11.018

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*(3), 524–539. doi:10.1037/0022-3514.81.3.524

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. doi:10.2307/271063

Reid, J. (1951). *The Reid Report*. (1951). Chicago, IL: John E. Reid & Associates.

Reinecke, J., Schmidt, P., & Ajzen, I. (1996). Application of the theory of planned behavior to adolescents' condom use: A panel study. *Journal of Applied Social Psychology*, *26*(9), 749–772. doi:10.1111/j.1559-1816.1996.tb01128.x

Richardson, M., & Abraham, C. (2009). Conscientiousness and achievement motivation predict performance. *European Journal of Personality*, *23*(7), 589–605. doi:10.1002/per.732

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*, *38*(2), 555–572. doi:10.5465/256693

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes, *13*(3), 279–301. doi:10.1016/0022-1031(77)90049-X

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634–644. doi:10.1037/0021-9010.83.4.634

Rossiter, J. (2009). *A Comparison of Social Desirability Bias among Four Widely Used Methods of Data Collection as Measured by the Impression Management Subscale of the Balance Inventory of Desirable Responding.* (Electronic Thesis or Dissertation). Retrieved from https://etd.ohiolink.edu/

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support?. *Human Resource Management Review*, *16*(2), 155–180. doi:10.1016/j.hrmr.2006.03.004

Rotundo, M., & Sackett, P. R. (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*(1), 66–80. doi:10.1037/0021-9010.87.1.66

Rust, J. (1999). The validity of the Giotto integrity test. *Personality and Individual Differences*, *27*(4), 755–768. doi:10.1016/S0191-8869(98)00277-3

Pedooem, R. (2007). *Relationship between HEXACO personality factors and task and contextual performance: Moderating effect of career advancement opportunity* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3305635)

Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, *65*(1), 70–89. doi:10.1177/0013164404268672

Peterson, M. H., Griffith, R. L., & Converse, P. D. (2009). Examining the role of applicant faking in hiring decisions: Percentage of fakers hired and hiring discrepancies in single- and multiple-predictor selection. *Journal of Business and Psychology*, *24*(4), 373–386. doi:10.1007/s10869-009-9121-5

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance*, *24*(3), 270–290. doi:10.1080/08959285.2011.580808

Peterson, M. H., Griffith, R. L., O'Connell, M. S., & Isaacson, J. A. (2008). Examining faking in real job applicants: A within-subjects investigation of score changes across applicant and research settings. In *R. L. Griffith & M. H. Peterson (Chairs), Examining faking*

using within-subjects designs and applicant data. Symposium conducted at the 23rd Annual Conference for the Society for Industrial and Organizational Psychology: San Francisco, CA. Retrieved from https://www.researchgate.net/profile/Matthew-Oconnell-7/publication/228479172_Examining_Faking_in_Real_Job_Applicants_A_Within-Subjects_Investigation_of_Score_Changes_Across_Applicant_and_Research_Settings/links/54e4de180cf22703d5bfa1eb/Examining-Faking-in-Real-Job-Applicants-A-Within-Subjects-Investigation-of-Score-Changes-Across-Applicant-and-Research-Settings.pdf [01.03.2021]

Pletzer, J. L., Bentvelzen, M., Oostrom, J. K., & De Vries, R. E. (2019). A meta-analysis of the relations between personality and workplace deviance: Big Five versus HEXACO. *Journal of Vocational Behavior*, *112*, 369–383. doi:10.1016/j.jvb.2019.04.004

Pletzer, J. L., Oostrom, J. K., Bentvelzen, M., & De Vries, R. E. (2020). Comparing domain- and facet-level relations of the HEXACO personality model with workplace deviance: A meta-analysis. *Personality and Individual Differences*, *152*, 109539. doi:10.1016/j.paid.2019.109539

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. doi:10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, *26*(3), 513–563. doi:10.1016/S0149-2063(00)00047-7

Sackett, P. R. (2011). Integrating and prioritizing theoretical perspectives on applicant faking of personality measures. *Human Performance*, *24*(4), 379–385. doi:10.1080/08959285.2011.597478

Sackett, P. R., Berry, C. M., Wiemann, S. A., & Laczo, R. M. (2006). Citizenship and counterproductive behavior: Clarifying relations between the two domains. *Human performance*, *19*(4), 441-464. doi.org/10.1207/s15327043hup1904_7

Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. *Personnel Psychology*, *42*(3), 491–529. doi:10.1111/j.1744-6570.1989.tb00666.x

Sackett, P. S. & DeVore, C. J. (2001). Counterproductive behaviour at work. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Visvesvaran (Eds.), *Handbook of industrial, work, and organizational psychology*, 1, 145–164. London: Sage.

Sackett, P. R., & Harris, M. M. (1984). Honesty testing for personnel selection: A review and critique. *Personnel Psychology*, *37*(2), 221–245. doi:10.1111/j.1744-6570.1984.tb01447.x

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, *102*(3), 254–273. doi:10.1037/apl0000151

Sackett, P. R., & Schmitt, N. (2012). On reconciling conflicting meta-analytic findings regarding integrity test validity. *The Journal of Applied Psychology, 97*(3), 550–556. doi:10.1037/a0028167

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty integrity, conscientiousness, dependability trustworthiness, and reliability for

personnel selection. *Personnel Psychology*, *49*(4), 787–829. doi:10.1111/j.1744-6570.1996.tb02450.x

Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, *82*(1), 30–43. doi:10.1037/0021-9010.82.1.30

Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, *10*(1-2), 117–125. doi:10.1111/1468-2389.00198

Salgado, J. F. (2017). Bandwidth-fidelity dilemma. In V. Zeigler-Hill, & T.K. Shackelford, (Eds.), *Encyclopedia of Personality and Individual Differences.* Springer International Publishing. doi:10.1007/978-3-319-28099-8_1280-1

Salgado, J. F., Moscoso, S., & Berges, A. (2013). Conscientiousness, its facets, and the prediction of job performance ratings: Evidence against the narrow measures. *International Journal of Selection and Assessment*, *21*(1), 74–84. doi:10.1111/ijsa.12018

Schleim, S. (2008). *Gedankenlesen. Pionierarbeit der Hirnforschung [Mind reading. Pioneering work in brain research]*. (1. Aufl.). Hannover: Heise.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. doi:10.1037/0033-2909.124.2.262

Schmidt, F. L., Shaffer, J. A., & Oh, I. S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, *61*(4), 827–868. doi:10.1111/j.1744-6570.2008.00132.x

Schneider, R. J., Hough, L. M., & Dunnette, M. D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior*, *17*(6), 639–655. doi:10.1002/(SICI)1099-1379(199611)17:6<639::AID-JOB3828>3.0.CO;2-9

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, *87*(2), 211–219. doi:10.1037/0021-9010.87.2.211

Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, *68*(4), 653–663. doi:10.1037/0021-9010.68.4.653

Snyman, J. H., Aamodt, M. G., Johnson, D. L., & Frantzve, J. (1991). Pre-employment paper-and-pencil honesty testing: A quantitative review. *Journal of Police and Criminal Psychology*, *7*(2), 11–25. doi:10.1007/BF02832823

Spector, P. E. (1997). The role of frustration in antisocial behavior at work. In R. A. Giacalone & J. Greenberg (Eds.), *Antisocial behavior in organizations* (p. 1–17). Sage Publications, Inc.

Spector, P. E., & Fox, S. (2002). An emotion-centered model of voluntary work behaviour. Some parallels between counterproductive work behaviour and organizational citizenship behavior. *Human Resource Management Review, 12*(2), 269–292. doi:10.1016/S1053-4822(02)00049-9

Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviors created equal?. *Journal of Vocational Behavior*, *68*(3), 446–460. doi:10.1016/j.jvb.2005.10.005

Staufenbiel, T., & Hartz, C. (2000). Organizational Citizenship Behavior: Entwicklung und erste Validierung eines Messinstruments [Organizational Citizenship Behavior: Development and validation of a measurement instrument]. *Diagnostica, 46*(2), 73–83. doi:10.1026/0012-1924.46.2.73

Striler, J., Shoss, M., & Jex, S. (2020). The relationship between stressors of temporary work and counterproductive work behaviour. *Stress and Health*. doi:10.1002/smi.2998

Stöber, J. (1999). Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität [The social desirability scale-17 (SDS-17): Development and first findings on reliability and validity]. *Diagnostica, 45*(4), 173–177. doi:10.1026//0012-1924.45.4.173

Sutherland, E. H. (1940). White-collar criminality. *American Sociological Review, 5*(1), 1–12. Retrieved from http://www.jstor.org/stable/2083937 [01.03.2021]

Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, *60*(4), 967–993. doi:10.1111/j.1744-6570.2007.00098.x

Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type. *Personality and Individual Differences, 52*(2), 195–201. doi:10.1016/j.paid.2011.10.017

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality–job performance relationship. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *24*(3), 335–356. doi:10.1002/job.191

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. doi:10.1037/bul0000217

Thomas, S. L., Grawitch, K. M., & Scandell, D. J. (2007). The double-rating method: A replication and extension. *Journal of Applied Social Psychology*, *37*(12), 2751–2763. doi:10.1111/j.1559-1816.2007.00279.x

Trevinyo-Rodríguez, R. N. (2007). Integrity: A systems theory classification. *Journal of Management History, 13*(1), 74–93. doi:10.1108/17511340710715188

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. doi:10.1037/0033-2909.133.5.859

Tsai, W.-C., Chen, H.-Y., & Chen, C.-C. (2012). Incremental validity of person-organization fit over the Big Five personality measures. *The Journal of Psychology, 146*(5), 485–509. doi.org/10.1080/00223980.2012.656154

U.S. Congress, Office of Technology Assessment (1983). *Scientific validity of polygraph testing* (OTA-TM-H-15). Washington, DC: U.S. Government Printing Office. Retrieved from: http://www.fas.org/sgp/othergov/polygraph/ota/index.html [01.03.2021]

Van Hooft, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301–316. doi:10.1037/a0025711

Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., & Putka, D. J. (2003). Applicant-incumbent differences on personality, integrity, and customer service measures. Poster presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012a). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, *97*(3), 499–530. doi.org/10.1037/a0021196

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012b). The critical role of the research question, inclusion criteria, and transparency in meta-analyses of integrity test research: A reply to Harris et al. (2012) and Ones, Viswesvaran, and Schmidt (2012). *The Journal of Applied Psychology, 97*(3), 543–542. doi.org/10.1037/a0026551

Van Iddekinge, C. H., Taylor, M. A., & Eidson, Jr, C. E. (2005). Broad versus narrow facets of integrity: Predictive validity and subgroup differences. *Human Performance*, *18*(2), 151–177. doi.org/10.1207/s15327043hup1802_3

Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, *81*(5), 525–531. doi:10.1037/0021-9010.81.5.525

Wanek, J. E. (1996). *The construct of integrity: Item level factor analysis of the dimensions underlying honesty testing and big-five measures of personality* (Doctoral dissertation). Institute: University of Minnesota.

Wanek, J. E., Sackett, P. R., & Ones, D. S. (2003). Towards an understanding of integrity test similarities and differences: An item-level analysis of seven tests. *Personnel Psychology*, *56*(4), 873–894. doi.org/10.1111/j.1744-6570.2003.tb00243.x

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, *19*(3), 188–202. doi:10.1016/j.hrmr.2009.03.007

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, *17*(3), 601–617. doi:10.1177/014920639101700305

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology, 29*(1), 64–77. doi:10.1080/1359432X.2019.1681401

Woolley, R. M., & Hakstian, A. R. (1992). An examination of the construct validity of personality-based and overt measures of integrity. *Educational and Psychological Measurement*, *52*(2), 475–489. doi:10.1177/0013164492052002024

Woolley, R. M., & Hakstian, A. R. (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*(1), 27–40. doi:10.1111/j.1468-2389.1993.tb00081.x

Yu, J. (2008). *A process model of applicant faking on overt integrity tests* (Doctoral dissertation), Texas A&M University. Retrieved from https://core.ac.uk/download/pdf/147131513.pdf

APPENDIX

Table A

Overview of international integrity tests

Test (Acronym[1])	Author	Year[2]	Type
Accutrac Evaluation System*[a]	B. R. Durbrow & B. H. Durbrow	1989	overt
ACM Attitude Evaluation	Magiera	1977	overt
AIMS ED3*[a]	Needham	1986	overt
Alienation Index Survey (A.I. Survey)	Cormark & Strand	1982	overt
Applicant Potential Inventory (API)*[ac]	National Computer Systems, Inc.	1997	overt
Applicant Review*	D. J. Cherrington & J. O. Cherrington	1993	overt
Applicant Risk Profiler (ARP)* b	Llobet	2001	overt
Attitudes to Honesty (ATH-2)*[ac]	Barrett	1987	overt & personality-based
California Psychological Inventory (CPI)[ac]	Gough	2005	personality-based
CIC Integrity Interview	Barke	1987	overt
Company Moral Questionnaire	Psychometric Behavioral Group	1979	overt
Compu-Scan	Terris	1986	overt
Counterproductive Behavior Index (CBI)*[a]	Goodstein & Lanyon	2003	overt
Employee Attitude Inventory (EAI-6)*[a]	London House, Inc.	1984	overt
Employee Integrity Index	Ryan & Sackett	1987	overt
Employee Perception Survey		1998	overt
Employee Reliability Index	J. Hogan & R. Hogan	1992	personality-based
Employee Reliability Inventory (ERI)*	Borofsky, Friedman, & Maddocks	1988	overt
Employee Screening Questionnaire (ESQ)*	Jackson	2002	personality-based
Employee Survey*	Bullard	1988	overt
Employee Attitude Inventory (EAI)	London House, Inc.	1984	overt
Employment Attitude Screening Evaluation (EASE II)*	Gelb & Rovner	1987	overt

Employment Inventory	Paajanen	1985	overt
Employment Productivity Index	Terris; London House, Inc.	1986	personality-based
Giotto*	Rust	1997	personality-based
Hogan Personnel Selection Series*	J. Hogan & R. Hogan	1995	personality-based
Hogan Personality Inventory–Reliability Scale (HPI-R)*[a]	Hogan	1984	personality-based
IntegriTEST[a]	Fine	2008	overt
Inwald Personality Inventory–Revised (IPI-R)*[a]	Inwald, Resko, & Favuzza	1996	personality-based
Inwald Survey 5–Revised (I IS5-R)[a]	Inwald		overt
Inventory of Job-Related Attitudes and Self-Assessments [Inventar berufsbezogener Einstellungen und Selbsteinschätzung] (IBES)*	Marcus	2006	overt & personality-based
Law Enforcement Applicant Inventory (LEAI)[c]	Pearson Reid London House		overt
Loss-Prevention Analysis*	D. J. Cherrington & J. O. Cherrington	1985	overt
Milby Profile*[a]	Miller & Bradly	1983	overt
Orion Survey*	Wilkerson	1985	overt
Orion Pre-Employment System			
PEAK Procedure*	-	1985	overt & personality-based
Pearson's Behavioral Construct*	Pearson	1977	overt & personality-based
PEOPLE Survey*	Hartnett & Teagle	1988	overt
Personnel Assessment Selection System (PASS)*[c]	Strand	1970	overt
Personnel Attitude Screening System (PASS III–Data Survey)*[a]	Cormack et al.	1987	overt
Personnel Decisions, Inc., Customer Service Inventory (PDI-CSI)*	Paajanen, Hansen, & McLellan	1993	overt & personality-based
Personnel Decisions, Inc., Employment Inventory (PDI-EI)*	Paajanen, Hansen, & McLellan	1993	overt & personality-based
Personnel Outlook Inventor (POI)*[a]	National Computer Systems, Inc.	1996	personality-based

Personnel Reaction Blank (PRB)*[a]	Gough, Arvey, & Bradley	2004	personality-based
Personnel Selection Inventory (PSI)*[ac]	Sackett & Harris	1992	overt
Personality Inventory to estimate Integrity [Persönlichkeitsinventar zur Integritätsabschätzung] (PIA)	Schuler & Fintrup	1999	overt
Phase II Profile*[c]	Lousig-Nont	1987	overt
Pre-employment Analysis Questionnaire	Gerhardt		overt
Preemployment Opinion Survey (POS)*	Harrelson, Paulson, & Yankee	1979	n.k.
PSC Survey A.D.T.*	A. L. Strand & M. L. Strand	1988	overt
PSC Survey L.T.*	A. L. Strand & M. L. Strand	1988	overt
Psychological Integrity Test [Psychologischer Integritätstest] (PIT)	Wilmer & Hoffmann	2006	overt
Reid Report*[ac]	Reid Psychological Systems; Brooks & Arnold	1988	overt
Reid Survey III*	Reid Psychological Systems	1988	overt
Security Aptitude Fitness Evaluation–Revised (SAFE-R)*[a]	Taccarino	1987	overt & personality-based
Sentry Survey*	Trego	1988	overt
Stanton Case Review	Klump	1987	overt.
Stanton Inventory*	Klump	1988	overt
Stanton Profile*	Harris	1987	overt
Stanton Survey* / Stanton Survey Phase II[a]	Klump	1988	overt
Station Employee Applicant Inventory (SEAI)*[c]	London House, Inc.	1986	overt
Station Manager Applicant Inventory (SMAI)*[c]	London House, Inc.	1986	overt
Step One Survey	Profiles International, Inc.	1998	overt
Substance abuse, Production loss, and Interpersonal problems Inventory	Caron	2003	overt
Trustworthiness Attitude Survey (T.A. Survey)*[ab]	Cormack & Strand	1970	overt
Tescor Survey*	Bullard	1986	overt
True Test*	Williams & Capps	1986	overt
Veracity Analysis Questionnaire	Bassett	1987	overt

(VAQ)*			
View Point (VP)*[a]	Gibson, Holcom, Stang, & Ruch	2003	overt & personality-based
Wilkerson Employee Input Survey*	Wilkerson	1986	overt
Wilkerson Preemployment Audit*	Wilkerson	1988	overt
Workplace Behavior Questionnaire	Ashton	1998	overt
Workplace Productivity Questionnaire	Nicol	1999	overt
WorkKeys Performance Assessment (WPA)*	ACT, Inc.	2007	overt & personality-based

Note. * = tests are included in the scale analysis, a = more than one version, b = more than one name, c = test is not available any more or status of publisher is unknown.

[1] = If there is an official test acronym, it is mentioned in parentheses. [2] = If there is more than one test version, the year of publication is referring to the latest version of the test.

Table B

Overview of experimental conditions referring to Hui's (2001) double-rating method, the method of indirect questioning, and to the control group (self-rating only)

Item and ratings		Double-rating method	Indirect questioning	Control group (self-rating only)
Dilemma situation of item		Nach einer Dienstreise zusammen mit drei seiner Arbeitskollegen macht Herr Müller mit seinen Kollegen die Dienstreiseabrechnung. Seine drei Kollegen wollen angeben, dass jeder selbst zum Zielort gefahren ist, obwohl Herr Müller und seine Arbeitskollegen gemeinsam in einem Auto angereist sind. Bei ungleichen Angaben in den Abrechnungen der Dienstreise könnten seine Kollegen in Schwierigkeiten kommen.		
Item		Bei einer gemeinsamen Dienstreiseabrechnung würde ich die gleichen Angaben wie meine mitgereisten Arbeitskollegen machen, auch wenn die Angaben nicht genau stimmen.		
Ratings	Other-rating	Wie sehr trifft diese Aussage auf andere Personen / Herrn Müller zu?	Wie sehr trifft diese Aussage auf andere Personen / Herrn Müller zu?	
	Self-rating	Wie sehr trifft diese Aussage auf Sie selbst zu?		Wie sehr trifft diese Aussage auf Sie selbst zu?
Rating scale		unzutreffend O	eher unzutreffend O	teils-teils O	eher zutreffend O	zutreffend O

Table C

Integrity-related items of the study on faking (Chapter 6)

Dilemma situation of item 1	Nach einer Dienstreise zusammen mit drei seiner Arbeitskollegen macht Herr Müller mit seinen Kollegen die Dienstreiseabrechnung. Seine drei Kollegen wollen angeben, dass jeder selbst zum Zielort gefahren ist, obwohl Herr Müller und seine Arbeitskollegen gemeinsam in einem Auto angereist sind. Bei ungleichen Angaben in den Abrechnungen der Dienstreise könnten seine Kollegen in Schwierigkeiten kommen.
Item 1	Bei einer gemeinsamen Dienstreiseabrechnung würde ich die gleichen Angaben wie meine mitgereisten Arbeitskollegen machen, auch wenn die Angaben nicht genau stimmen.
Dilemma situation of item 2	Frau Meyer hat seit Längerem geplant, heute eher von der Arbeit zu gehen, um sich mit einer Freundin zu treffen. Im Laufe des Tages bemerkt Frau Meyer, dass sie heute viel zu erledigen hat und deshalb die Verabredung nicht einhalten kann. Da der Akku ihres Handys fast leer ist, überlegt sie, die Verabredung von ihrem Bürotelefon aus abzusagen, damit ihre Freundin nicht unnötig warten muss.
Item 2	Private Telefonate würde ich nicht von der Arbeit aus erledigen, auch wenn ich eine dringende Nachricht für einen Freund habe.
Dilemma situation of item 3	Es ist spät am Abend und Herr Schmidt ist noch auf der Arbeit. Plötzlich klingelt sein Telefon und ein Freund meldet sich. Dieser ist verzweifelt, weil er dringend die Kopie eines wichtigen Dokuments benötigt. Der Freund erklärt Herrn Schmidt, dass es Probleme mit seinem Drucker gab und die Kopierläden der Stadt schon geschlossen sind. Er bittet ihn daher, das wichtige Dokument für ihn zu kopieren.
Item 3	Auch wenn ein Freund dringend eine Kopie braucht, würde ich ihm diese nicht auf der Arbeit anfertigen.
Dilemma situation of item 4	Frau Becker ist in Urlaubsstimmung. Morgen früh ist es endlich soweit und es geht in den wohlverdienten Urlaub. Frau Becker fragt sich, wie das Wetter wohl gerade in ihrem Urlaubsort ist. Sie überlegt, sich im Internet zu informieren, was sie die kommenden zwei Wochen klimatechnisch erwarten wird, schließlich geht es heute Abend ans Packen und sie muss wissen, für welche Wetterverhältnisse sie sich rüsten muss.
Item 4	Ich würde während der Arbeitszeit keine privaten Angelegenheiten erledigen.
Dilemma situation of item 5	Frau Schulz ist im Stress. Ihr Geburtstag steht vor der Tür und bis jetzt hat sie noch keine Zeit gefunden, sich um die Planung ihrer Geburtstagsparty zu kümmern. Dabei sind es nur noch acht Tage und sie hat noch nicht einmal die Räumlichkeiten für ihre Feier reserviert. Frau Schulz überlegt, ob sie nicht gleich eine Reservierungsanfrage per E-Mail versenden sollte, dann wäre zumindest das Wichtigste erledigt.

Item 5	Ich würde während der Arbeitszeit E-Mails verfassen und versenden, die einen privaten Inhalt enthalten.
Dilemma situation of item 6	Aufgrund der guten Auftragslage wächst Herrn Neumann die Arbeit gerade über den Kopf. Das Arbeitspensum ist kaum noch zu bewältigen. Aufgaben, die schon längst hätten erledigt sein sollen, bleiben einfach liegen. Als der Vorgesetzte von Herrn Neumann nachfragt, ob er die Broschüren für die Messe fristgerecht gestern bestellt hat, überlegt Herr Neumann, dies zu bejahen, obwohl er es bisher nicht geschafft hat, die Broschüren zu ordern. Er könnte die Bestellung unverzüglich nachholen, so dass höchstwahrscheinlich noch der ursprüngliche Liefertermin eingehalten werden könnte.
Item 6	Ich würde meinen Vorgesetzten belügen, um einen Fehler meinerseits zu vertuschen.