
Bewährung von Integritätstests in der Praxis – Studienergebnisse aus dem öffentlichen Sektor

2

Andreas Gourmelon

2.1 Korruptionsprävention durch Integritätstests

Integrität ist ein wichtiger Aspekt der charakterlichen Eignung von Beschäftigten im öffentlichen Sektor. Korruption gefährdet den Rechtsstaat, indem sie dessen tragende Prinzipien, wie den Grundsatz der Rechtsstaatlichkeit oder den Grundsatz der Gleichbehandlung, außer Kraft setzt. Korruptionsprävention ist in den letzten zehn Jahren auch in der öffentlichen Verwaltung zunehmend ein Thema geworden. Die entsprechenden Stellen haben mit Dienstverpflichtungen, Dienstvereinbarungen, Anweisungen oder eigenen Richtlinien sukzessive reagiert. Darüber hinaus hat in den letzten Jahren eine Vielzahl von Bundes-, Landes- und Kommunalverwaltungen spezielle Organisationseinheiten eingerichtet, die für Korruptionsprävention und interne Revision zuständig sind (Sorgatz 2012, S. 5). Dabei bleibt ein Punkt jedoch in der Regel ausgeklammert, nämlich die Frage, wie man bereits bei der Personalauswahl auf das Problem reagieren kann. Insbesondere im Zusammenhang mit der Prävention von Korruption gewinnt die Überprüfung der Integrität im Rahmen der Personalauswahl an Bedeutung (Gourmelon et al. 2013, S. 1). Dementsprechend forderte die Freie und Hansestadt Hamburg (2013):

Bei der Rekrutierung von Personal sind Fragen des „moralischen Bewusstseins“ und der charakterlichen Eignung stärker zu berücksichtigen. Das gilt sowohl im Hinblick auf den Eintritt in den öffentlichen Dienst als auch für bestimmte (korruptionsgefährdete) Tätigkeitsbereiche oder Führungsfunktionen. Zur besseren Beurteilung der Integrität einer Person sollen entsprechend geeignete Instrumente entwickelt werden.

A. Gourmelon (✉)
FH für öffentliche Verwaltung NRW, Gelsenkirchen, Deutschland
E-Mail: andreas.gourmelon@fhoev.nrw.de

Für die Korruptionsprävention kann auf sogenannte Integrity-Tests zurückgegriffen werden. Integrity-Tests messen das Persönlichkeitskonstrukt Integrität und sollen die Vertrauenswürdigkeit einer Person erfassen. Sie werden mit dem Ziel konstruiert, das Kriterium kontraproduktives Verhalten am Arbeitsplatz vorherzusagen (Zettler und Solga 2008, S. 98). Kontraproduktives Verhalten umfasst neben Korruption auch Sabotage, Diebstahl u. v. m. So existieren vielfältige Ansätze in Form von Integrity- und Compliance-Systemen, die als Präventionsmaßnahmen die Integrität im Unternehmen fördern sollen. Obwohl diese ihre Wirkungen erzielen, wird mit Integritätstests nicht der Situationscharakter betont, sondern diese Tests legen den Schwerpunkt auf eine personalistische Erklärung kontraproduktiven Verhaltens, den dispositiven Ursachen in der Person des einzelnen Mitarbeiters (Marcus 2000, S. 342). Integre Menschen zeigen eine höhere berufliche Leistung und neigen weniger zu rechtlich und ethisch abweichendem Verhalten. Integritätstests haben in den USA eine über 60 Jahre alte Tradition (Marcus 2006, S. 9). Dort werden sie mittlerweile bei über 50 % der Auswahlverfahren durchgeführt. In Deutschland ist die Messung der Integrität der Bewerber noch nicht weitverbreitet. Die ersten Tests wurden Ende der 1990er-Jahre angewendet, die Einsatzhäufigkeit liegt bei unter einem Prozent (Frintrup et al. 2004, S. 58 ff.). Nach ersten Recherchen gibt es keine veröffentlichten Erfahrungsberichte zum deutschsprachigen Integritätstest „Inventar berufsbezogener Einstellungen und Selbsteinschätzungen“ (IBES). Deutsche Unternehmen halten sich bedeckt, wenn sie nach der Anwendung des IBES gefragt werden (Koch 2013).

In der interdisziplinären Forschungsarbeit von Gourmelon et al. (2013), aus der im Folgenden ausgewählte Ergebnisse vorgestellt werden, wird der Frage nachgegangen, ob mit Integritätstests Korruption und anderes organisationsschädigendes/kontraproduktives Handeln prognostiziert werden kann und der Einsatz von Integritätstests im Rahmen der Personalauswahl des öffentlichen Sektors aus rechtlichen Gründen möglich ist.

Zur Heranführung an die Forschungsergebnisse wird zunächst die historische Entwicklung von Integritätstests aufgezeigt und erläutert, was diese Tests eigentlich genau messen. Zudem werden empirische Befunde zur Aussagekraft dieser Tests dargestellt. Anschließend wird am Beispiel des IBES erläutert, wie diese Tests in der Praxis durchgeführt werden.

Des Weiteren werden ausgewählte Befunde zur Qualität des IBES, insbesondere zur Vorhersagegüte, dargelegt und schließlich Empfehlungen für die Praxis der Personalauswahl gegeben.

2.2 Beschreibung von Integritätstests

Nach Ones und Viswesvaran (2001) sind Integritätstests „... paper and pencil tests specifically developed to assess dependability, integrity and honesty of applicants thereby facilitating prediction of theft and future on-the-job dishonest behaviours“ (S. 32). Integritätstests werden in den USA seit über 60 Jahren eingesetzt. Der Begriff steht für zwei historisch

und konzeptionell klar unterscheidbare Verfahrensgruppen und unabhängig voneinander entwickelte Forschungsstränge, für die sich die Begriffe „*overt*“ (offenkundig) bzw. „*personality-based*“ (persönlichkeitsorientiert) durchgesetzt haben (Marcus 2006, S. 9).

2.2.1 Historische Entwicklung

Integritätstests unterschiedlicher Arten waren unabhängig voneinander entstanden, lange bevor sie von der Fachöffentlichkeit als Verfahrensgruppe wahrgenommen und klassifiziert wurden. Diese Verfahren messen „... eine im Detail sehr heterogene Mischung unterschiedlicher Einstellungen, Persönlichkeitseigenschaften und biografisch manifestierter Verhaltenstendenzen, die jedoch alle empirisch mit dem interessierenden Kriterium der Kontraproduktivität zusammenhängen“ (Marcus 2006, S. 10). Integritätstests wurden aber nicht entwickelt, um das Konstrukt der Integrität zu messen. Die Bezeichnung „Integrität“ wurde nachträglich für bereits existierende Verfahren gewählt (Marcus 2006, S. 10).

Wie viele eignungsdiagnostische Verfahren entstammen Integritätstests ursprünglich dem militärischen Bereich. Während des Zweiten Weltkrieges entwickelte der amerikanische Armeepsychologe G. L. Betts ein Instrument, mit dessen Hilfe sich Rekruten mit kriminellem Hintergrund identifizieren lassen konnten. Es entstand ein 67-Item-Fragebogen, der neben Eingeständnissen abweichenden Verhaltens (z. B. „Wie oft haben Sie Sachen gestohlen, bevor sie 12 Jahre alt waren?“) auch Einstellungen (z. B. „Wie weit kann man den meisten Menschen trauen?“) berücksichtigte. Dieser Test, *Biographical Case History (BCH)*, stellt den ersten, direkt für Auswahlzwecke konzipierten Test zur Prognose devianten Verhaltens dar. Der *Biographical Case History* gilt als Vorreiter für den einstellungsorientierten Typus von Integritätstests (Marcus et al. 1997, S. 3).

Unabhängig von der Arbeit Betts' begann Ende der 1940er-Jahre der Jurist John Reid einen Test zu entwickeln, der unter dem Namen *Reid Report* 1951 erstmals vorgestellt wurde. Dieser Test, der zum damaligen Zeitpunkt den Einsatz des Polygrafen optimieren sollte, untersucht die Einstellungen und Wertehaltungen des Bewerbers hinsichtlich devianten Verhaltens. Der Reid Report wird wegen seiner direkten Fragestellung auch als „*Overt*“-Typ (offenkundig) bezeichnet. Er zählt heute, nach mehrfachen Revisionen, zu den verbreitetsten Integritätstests. Beiden Arten von Integritätstests ist gemeinsam, dass die Item-Selektion vorwiegend in Kontrastgruppen erfolgte, d. h., es wurden aus einem großen Item-Pool diejenigen Fragen ausgewählt, die am besten zwischen beispielsweise einer Gruppe Strafgefangener und einer Normalpopulation trennten (Marcus et al. 1997, S. 3).

Daran angelehnt entstanden viele Verfahren dieser Art. Im Jahr 1989 zählte die Monografie von O'Bannon et al. bereits 43 derartige Tests, und einige Jahre darauf konnten für eine Metaanalyse¹ (Ones et al. 1993) weit über 600 Validierungsstudien zu diesen Verfahren herangezogen werden (vgl. Marcus 2000, S. 8).

¹Mit Metaanalysen werden die Ergebnisse verschiedener empirischer Studien analysiert und zusammengefasst.

In Deutschland existieren zwei Anbieter von Integritätstests, die diese ausschließlich an professionelle Endkunden vertreiben, die ihrerseits umfangreiche Schutzverpflichtungen eingehen (Tenzer 2005, S. 49). Seit dem Jahr 2000 ist das „Persönlichkeitsinventar zur Integritätsabschätzung“ (kurz: PIA) der „S & F Personalpsychologie“ auf dem Markt, es folgte der „Psychologische Integritätstest“ (kurz: PIT) des „Team Psychologie & Sicherheit“. Der genaue Aufbau der Tests oder die Testfragen werden nach Frintrup (2005) nicht veröffentlicht, „... weil sich die Inhalte und der Auswertungs- und Lösungsalgorithmus schnell herumsprechen würden“ (Tenzer 2005, S. 49). Im Jahr 2006 kam das „Inventar berufsbezogener Einstellungen und Selbsteinschätzungen“ (Marcus 2006) auf den Markt. Es ist das erste in deutscher Sprache veröffentlichte Verfahren vom Typus eines Integritätstests.

2.2.2 Messung der Integrität

Integrität bezeichnet allgemein das Handeln nach moralischen Standards und Prinzipien, dazu zählen beispielsweise wahrheitsgemäße Aussagen, das Einhalten von expliziten und impliziten Versprechen und die Übereinstimmung von Worten und Taten (Friedrich 2003, S. 119 ff.). Mit dem Begriff Integrität, der in der Philosophie und Ethik als Einheitlichkeit zwischen Verhalten und eigenen moralischen Maßstäben, mit den Dimensionen Selbsttreue, Rechtschaffenheit und Unbestechlichkeit (Pollmann 2005, S. 77 ff.), definiert wird, stimmen die in Integritätstests erfassten Dispositionen aber „... höchstens zufällig und ausnahmsweise überein“ (Marcus 2006, S. 10). Marcus entscheidet sich so für einen theoretischen Zugang zum Begriff der Integrität, der ohne moralische Kategorien auskommt. Kontraproduktives Verhalten kann als Gegenpol zu Integrität bezeichnet werden (Hoffmann et al. 2006, S. 59 ff.). In Bezug auf das Verhalten am Arbeitsplatz kann Integrität folglich als Handeln im Sinn und zum Wohl der Organisation definiert werden.

Der Begriff des kontraproduktiven Verhaltens wird oft synonym zum Begriff der Devianz verwendet. Devianz bezeichnet die Abweichung von einer Norm. Joas (2001, S. 184) definiert Abweichung als „... jede Handlung, von der angenommen wird, dass sie eine allgemein geltende Norm einer Gesellschaft oder einer bestimmten Gruppe dieser Gesellschaft verletzt“. Nach Mussel (2003) ist der Begriff der Devianz weiter gefasst als der des kontraproduktiven Verhaltens, da Devianz sich nicht nur auf arbeitsrelevantes Verhalten bezieht. Kontraproduktives Verhalten ist die Organisation und die Mitarbeiter schädigendes Verhalten. Dazu gehören aggressives Verhalten gegenüber Kollegen oder Dingen am Arbeitsplatz, Korruption, Amtsmissbrauch, Mitarbeiterdiebstahl, Absentismus oder Substanzmissbrauch. Der Schaden, der der deutschen Volkswirtschaft dabei entsteht, wird auf jährlich mehrere Milliarden Euro geschätzt. Nach Marcus (2000, S. 1) ist es ein offenes Geheimnis, dass „... Mitarbeiter sich gelegentlich an der Kasse oder dem Warenregal vergreifen, kurze krankheitsbedingte Fehlzeiten vorzugsweise am Wochenende herum auftreten oder in manchem Schreibtisch oder Spind hochprozentiger Trost für den

anstrengenden Arbeitsalltag lagert“. Ein offenes Geheimnis insofern, als dass sowohl die Personalverantwortlichen als auch die Wissenschaft eine Haltung des „Nichtdranrührens“ einnehmen (Marcus 2000, S. 1), respektive die Ursachen vorwiegend in situativen Umständen gesucht werden bzw. in Personalvariablen, die man sich ihrerseits als vorwiegend durch die Situation bestimmt vorstellt (Arbeitszufriedenheit). So existieren viele unterschiedliche Managementsysteme, die genau dort ansetzen, also die versuchen, integrires Handeln der Mitarbeiter durch Arbeitsplatzgestaltung zu beeinflussen. Gleichwohl die bedeutende Rolle äußerer Gegebenheiten nicht bestritten wird (vgl. Marcus 2006, S. 106 ff.), legen Integritätstests bzw. deren Konzeption den Schwerpunkt auf eine personalistische Erklärung kontraproduktiven Verhaltens, den dispositiven Ursachen in der Person des einzelnen Mitarbeiters (Marcus 2000, S. 342). Nach Marcus (2006, S. 15) deuten geringe Werte in Integritätstests auf eine erhöhte Attraktivität kontraproduktiven Verhaltens hin. Er beschreibt die Ursache der erhöhten Attraktivität kontraproduktiven Verhaltens in einem dispositionalen Mangel an Unterdrückung gegenüber Impulshandlungen. Diese Annahme konnten Marcus et al. (2002, S. 18 ff.) in einer Untersuchung über die Prädiktoren für Kontraproduktivität bestätigen. Demnach war die Selbstkontrolle der beste Prädiktor für Kontraproduktivität. Mit Integritätstests werden so Merkmale erfasst, die sich teils den motivationalen Dispositionen (Neigungen) und teils dem Bereich der internalen Kontrolle zuordnen lassen. Mit der Erfassung dieser Merkmalsausprägungen können Personen sowohl hinsichtlich ihrer zu erwartenden Arbeitsleistung als auch hinsichtlich der Wahrscheinlichkeit, mit der sie deviante Verhaltensweisen zeigen, unterschieden werden. Mit Integritätstests werden aber keine Merkmale der Situation erfasst. Insofern leisten sie einen Beitrag zur Verhaltensprognose, bieten aber keine vollständige Verhaltensklärung (Marcus 2006, S. 15). Nach Marcus (2006) kommt es bei der Auslegung des Begriffes kontraproduktiven Verhaltens zunächst nicht darauf an, dass der Organisation tatsächlich ein Schaden entsteht. Entscheidend ist vielmehr zunächst, dass aus dem Verhalten ein negatives Resultat entstehen kann und sich diese Möglichkeit dem Handelnden ohne besondere Vorkenntnisse erschließt (Marcus 2006, S. 81). Kontraproduktives Verhalten kann als Gegenpol zu Integrität bezeichnet werden (Hoffmann et al., 2006, S. 59 ff). In Bezug auf das Verhalten am Arbeitsplatz kann Integrität folglich als Handeln im Sinn und zum Wohl der Organisation definiert werden.

Nach metaanalytischen Befunden (Ones et al. 1993) konnten in 23 Studien mit 7550 Teilnehmern Zusammenhänge von Arbeitsleistung mit den in Integritätstests erzielten Werten von $r = 0,41^2$ nachgewiesen werden. Der Zusammenhang mit kontraproduktivem Verhalten liegt bei 0,47. Van Iddekinge et al. (2012, S. 399 ff.) berichten von einem Korrelationskoeffizienten in Höhe von $r = 0,42$ für die Prognose von kontraproduktiven Handlungen, die mit Selbstauskünften erfasst wurden.

²Die Korrelation ist der statistische Zusammenhang zweier veränderlicher Größen. Die Höhe der Korrelation wird mit dem Korrelationskoeffizienten r gemessen. $r = 0$ bedeutet, dass kein Zusammenhang, $r = 1$, dass ein maximaler Zusammenhang besteht.

2.2.3 Funktionsweise von Integritätstests am Beispiel IBES

Mit dem „Inventar berufsbezogener Einstellungen und Selbsteinschätzungen“ (IBES) liegt erstmals ein in deutscher Sprache publiziertes Verfahren vom Typus der Integritätstests vor. Das IBES wurde explizit berufsbezogen konzipiert und beschränkt sich demzufolge auf die Erfassung von Merkmalen, deren berufliche Relevanz empirisch bestätigt ist. Es kann als Einzel oder Gruppentest eingesetzt werden und erfordert circa 25 min Bearbeitungszeit. Es beschränkt sich nicht auf bestimmte Branchen, Hierarchien oder Berufsgruppen. Außerdem ist es unabhängig vom Qualifikationsniveau der Teilnehmer. Das IBES ist ein Selbstberichtsverfahren, bestehend aus 115 Items in neun Subskalen, die sich teils der einstellungs-, teils der eigenschaftsorientierten Variante von Integritätstests zuordnen lassen. Zum ersten Teil gehören die Facetten Vertrauen, Verbreitung unerwünschten Verhaltens, Nichtrationalisierung und Verhaltensabsichten, zum zweiten Teil Gelassenheit/Selbstwertgefühl, Zuverlässigkeit/Voraussicht, Vorsicht, Zurückhaltung und Konfliktmeidung. Der Gesamtwert des IBES, und damit die Summe der Ausprägungen dieser verschiedenen Merkmale, die sich als empirisch relevant für das Kriterium kontraproduktiven Verhaltens erwiesen haben, liefert eine Aussage über die relative Wahrscheinlichkeit des späteren Auftretens kontraproduktiver Handlungen. Die hinter dem Gesamtwert stehenden Konstrukte werden beim IBES offengelegt und einer separaten Interpretation zugänglich gemacht. Gleichwohl soll, was der Testautor mehrfach anmerkt, bei Personalentscheidungen der Gesamtwert berücksichtigt werden und dieser nicht als alleiniges Instrument für eine Personalauswahlentscheidung herangezogen werden (Marcus 2006, S. 29). Die Wahrscheinlichkeit kontraproduktiver Handlungsweisen eines Bewerbers ist umso niedriger,

- je mehr dieser andere Menschen für vertrauenswürdig hält;
- je weniger er davon ausgeht, dass Regelverletzungen im Berufsleben weitverbreitet sind;
- je weniger er dazu neigt, Regelverstöße durch „vernünftige“ Gründe zu rechtfertigen;
- je weniger er sich gedanklich mit Regelverstößen auseinandersetzt oder diese gar plant;
- je gelassener und selbstsicherer er ist;
- je zuverlässiger und kontrollierter er ist;
- je mehr er Risiken und Aufregung meidet;
- je ruhiger, umgänglicher und bescheidener er ist;
- je mehr er dazu neigt, Konflikte zu vermeiden und Harmonie anzustreben.

Zur Auswertung werden Summenblätter und Schablonen herangezogen und die auf den Schablonen befindlichen, für jedes Item unterhalb der Antwortskalen stehenden Rohwerte in das Summenblatt eingetragen. Bei der Übertragung der Rohwerte in der Schablone auf das Summenblatt muss die unterschiedliche Polung (Zustimmungsskala vom Likert-Typ: starke Ablehnung – Ablehnung – neutral – Zustimmung – starke Zustimmung) verschiedener Items berücksichtigt werden. Im Anschluss daran werden die Summenrohwerte je

Subskala berechnet, indem die Einzelwerte jeder Spalte addiert und in das Feld „Summe“ am Ende jeder Spalte eingetragen werden. Diese neun Summenwerte werden dann in die letzte Spalte übertragen und wiederum aufaddiert. Dies ergibt den Rohwert für das Gesamtergebnis im IBES. Zu den Summenrohwerten werden im letzten Schritt die zugehörigen Normwerte in den Normtabellen nachgeschlagen und in die entsprechenden Felder des Summenblatts eingetragen. In den Normtabellen sind zu jedem Rohwert zwei Normwerte aufgeführt, Prozentränge und Standard-Nine (STANINE)-Werte, die unterschiedlich zu interpretieren sind (Marcus 2006, S. 31). Normiert wurde der Test in den Jahren 2002/2003 an insgesamt 332 externen Bewerbern aus unterschiedlichen Berufsgruppen und Branchen. Die Anwendung des IBES ist einfach und – da als Gruppentest durchführbar – ökonomisch.

Das Gütekriterium der Objektivität wird durch Beobachtung der Testdurchführung, der neutralen Haltung des Testleiters und der standardisierten Auswertung des Tests erfüllt. Für die Reliabilität/Messgenauigkeit liegen in fast allen Fällen der Erhebungen zufriedenstellende Angaben für die interne Konsistenz und Retest-Reliabilität vor (Marcus 2006, S. 39). Das IBES weist ein Cronbachs Alpha von 0,91 und 0,94 für den Gesamtwert auf. „Die interne Konsistenz des einstellungs- (alpha = .89 bis .92) und eigenschaftsorientierten (alpha = .83 und .88) Teils des IBES liegt in allen Stichproben oberhalb der für vergleichbare amerikanische Verfahren berichteten Mittelwerte (alpha = .83 bzw. .72; Ones et al. 1993)“ (Marcus 2007, S. 130). Die Retest-Reliabilität über einen Zeitraum von fünf Monaten beträgt $r = 0,89$. Mit einem r von 0,02 und 0,07 kann ebenso die Unabhängigkeit zur allgemeinen Intelligenz als gegeben angesehen werden. Dies ist insofern relevant, als dass Integritätstests die valideste Zusatzinformation über den besten Einzelprädiktor beruflicher Leistung, die allgemeine Intelligenz, liefern (Schmidt und Hunter 1998). Alle Korrelationen mit dem Gesamtwert des IBES liegen zwischen 0,46 und 0,55 hinsichtlich des Kriteriums kontraproduktiven Verhaltens. Ebenfalls mit großer Konsistenz korreliert das IBES mit fremd beurteilten beruflichen und akademischen Leistungen, jeweils mit 0,20, sowie mit beruflicher Zufriedenheit mit 0,30 (Marcus 2006, S. 5).

In Anlehnung an erste Befunde ist davon auszugehen, dass der Einsatz des IBES in der Personalauswahl zu verbesserten Personalentscheidungen führen kann:

Für das IBES sind bislang kaum Studien zur Prognosegüte bekannt. Hanke (2011) überprüfte die Geeignetheit des IBES zur Vorhersage der Arbeitsleistungen von Auszubildenden eines Versicherungsunternehmens. Es zeigten sich mäßige, teils signifikante statistische Zusammenhänge zwischen dem IBES-Gesamtwert und den Ergebnissen dienstlicher Beurteilungen (Hanke 2011, S. 125). Mit dem Zusammenhang zwischen IBES-Testergebnissen und kontraproduktiven Handlungen befassten sich Marcus (2006), Marcus und Wagner (2007), Marcus et al. (2007) sowie Marcus et al. (o. J.).

Exemplarisch werden nachfolgend die Vorgehensweise und die Ergebnisse der Studie von Marcus (2006) dargestellt: 168 Berufstätige im Einzelhandel und der Industrie bearbeiteten zuerst das IBES (sie sollten wie in einer Bewerbungssituation antworten),

und unmittelbar anschließend gaben sie unter anonymen Bedingungen an, wie oft sie im vergangenen Jahr kontraproduktive Handlungen ausgeführt hatten. Hierfür hatten die Befragten eine Liste mit 50 kontraproduktiven Handlungsweisen vorliegen. Der statistische Zusammenhang zwischen dem IBES-Gesamtwert und dem Ausmaß kontraproduktiver Handlungen beträgt $r = 0,46$ (Marcus 2006, S. 50), in Tab. 2.1 sind weitere Ergebnisse der Studie aufgelistet.

Aus Tab. 2.1 ist beispielsweise ersichtlich, dass sich das Risiko des Auftretens eines Kündigungsgrundes bei einem zukünftigen Mitarbeiter beim Anwenden des IBES im Rahmen der Personalauswahl auf ein Siebtel reduzieren ließe (0,07 zu 0,49) (Gourmelon et al. 2013, S. 8).

In Bezug auf den Einsatz von Persönlichkeitstests in der Personalauswahl finden sich insbesondere zwei Umstände, die Personalverantwortliche von dem Einsatz dieser Verfahren abhalten. Ein Kritikpunkt ist, dass Personen die Ergebnisse, aufgrund einer Tendenz zu „sozial erwünschtem“ Antwortverhalten, verfälschen können. Ein weiterer ist die etwaige fehlende soziale Validität, die soziale Akzeptanz aufseiten der Bewerber.

In der Literatur finden sich verschiedene theoretische Auffassungen zum Themengebiet der Verfälschungstendenzen. In Bezug auf das IBES führt Marcus (2006, S. 29) aus, dass Bewerber im Mittel höhere Werte auf sozial bzw. beruflich erwünschten Dimensionen erreichen als Teilnehmer anonymer Untersuchungen und dass diese Dimensionen bei Bewerbern untereinander statistisch stärker zusammenhängen. Diesem Umstand trägt der Testautor aber Rechnung, indem er die theoretischen Beziehungen zwischen den Skalen bei Nichtbewerbern untersuchte, das IBES aber an Bewerbern normiert wurde. Die Tatsache, dass Antworten im Sinne sozialer Erwünschtheit gegeben werden, hat nach

Tab. 2.1 Ausmaß kontraproduktiven Verhaltens in Abhängigkeit vom IBES-Gesamtwert. (Nach Marcus 2006, S. 35)

| IBES-Gesamtwert ist ... | Von den Befragten hätten eine Abmahnung erhalten ...% | Von den Befragten wären gekündigt worden ...% | Mittelwert Anzahl Abmahnungsgründe | Mittelwert Anzahl Kündigungsgründe |
|--------------------------------|---|---|------------------------------------|------------------------------------|
| Deutlich unterdurchschnittlich | 100 | 39,3 | 10,29 | 1,57 |
| Knapp unterdurchschnittlich | 85,7 | 28,1 | 6,66 | 0,53 |
| Knapp überdurchschnittlich | 76,4 | 10,9 | 3,11 | 0,16 |
| Deutlich überdurchschnittlich | 53,6 | 7,1 | 1,96 | 0,07 |
| Insgesamt (N = 168) | 78,9 | 20,3 | 5,23 | 0,49 |

Marcus (2006) keinen Einfluss auf die kriteriumsbezogene Validität. Auch auf den Einsatz sogenannter Erwünschtheitsskalen, die solche Tendenzen messbar und nachträglich korrigierbar machen sollen, hat der Testautor, aufgrund der metaanalytisch gewonnenen Erkenntnis, dass die Auspartialisierung solcher Tendenzen keinen Effekt auf die kriteriumsbezogene Validität hat (Ones und Viswesvaran 2001), verzichtet (Marcus 2006, S. 46).

Ein weiteres wesentliches Gütekriterium ist die Akzeptanz des Verfahrens seitens der Bewerber. Reinhold (2007) führt hierzu aus, dass es Personalverantwortlichen schwerfalle, diese Tests in der Personalauswahl einzuführen, da solche Tests hierzulande, im Gegensatz zu den USA, mit negativen Aspekten – wie „Aushorchen“ – in Verbindung gebracht würden. Marcus (2006, S. 54) untersuchte die soziale Akzeptanz bei Nichtbewerbern, die folglich keine Konsequenzen bei eventueller negativer Rückmeldung zu befürchten hatten, und fand heraus, dass hier von einer neutralen Haltung der Probanden gegenüber dem IBES ausgegangen werden kann, was eine vernünftige Basis für einen, durch Voreinstellungen und Reaktionen unbelasteten, Einsatz in der Praxis ermögliche.

2.3 Integritätstests bei der Auswahl des Beamtennachwuchses

Nachfolgend werden ausgewählte Ergebnisse eines Forschungsprojekts dargestellt, mit dem geprüft werden sollte, ob Integritätstests (hier das IBES) nutzbringend in der Personalauswahl des öffentlichen Sektors verwendet werden können. Ausführliche Darstellungen finden sich bei Gourmelon (2016) und Michaelis (2015).

In einem ersten Schritt wurde im Herbst 2012 Studierenden der Fachhochschule für öffentliche Verwaltung Nordrhein-Westfalen (FHöV NRW) das IBES zur Bearbeitung vorgegeben. Studierende der FHöV NRW befinden sich in der Regel bereits in einem Beamten-Dienstverhältnis, das Studium stellt für sie eine dienstliche Aufgabe dar.

Im Jahr 2015 wurden dieselben Studierenden erneut befragt. Bei dieser nochmaligen freiwilligen und anonymen schriftlichen Befragung wurde erhoben, inwieweit die Studierenden während des Studiums kontraproduktive Handlungen ausgeübt haben. Zudem wurde nochmals der IBES vorgegeben.

In einem dritten Schritt wurde der statistische Zusammenhang zwischen dem IBES-Gesamtwert und dem Ausmaß kontraproduktiver Handlungen berechnet.

Insgesamt konnten die Daten von 92 Studierenden verwendet werden. 80,4 % der Studierenden ließen sich zu einer Tätigkeit in der Kommunalverwaltung, 19,6 % zu einer Tätigkeit im Polizeivollzugsdienst ausbilden. Ein Drittel der Befragten war männlichen, zwei Drittel waren weiblichen Geschlechts. An einem Gymnasium hatten 75,3 %, an einer anderen Schule 24,7 % der befragten Studierenden die Hochschulreife erworben. Die Abitur-Durchschnittsnote der ehemaligen Gymnasiasten betrug 2,65 (Varianz: 0,49). Erfahrungen an einer Hochschule vor dem Studium an der FHöV NRW hatten 22,8 %

der Befragten, Erfahrungen in einer Berufstätigkeit im öffentlichen Sektor 26,1 %, Erfahrungen in einer Berufstätigkeit außerhalb des öffentlichen Sektors 17,4 % der Studierenden. Knapp 49 % der befragten Studierenden hatten vorab keine Erfahrungen im Berufsleben oder an einer Hochschule erworben. Angesichts der Ergebnisse eines t-Tests müssen die Unterschiede in den IBES-Gesamtwerten 2012 und 2015 als zufällig gelten. Männer und Frauen sowie Polizei- und Kommunalbeamte unterscheiden sich im Mittelwert der IBES-Gesamtwerte 2012 und 2015 nur zufällig (geprüft mit t-Test). Der statistische Zusammenhang zwischen Abiturnote und dem IBES-Gesamtwert beträgt erwartungsgemäß bei der Stichprobe der FHöV-Studierenden im Jahr 2012 $r = 0,17$ und im Jahr 2015 $r = 0,23$.

Für die Qualität eines Testverfahrens ist wesentlich, dass zu zwei verschiedenen Zeitpunkten bei einzelnen Personen ähnliche Messergebnisse erzielt werden. Die Bestimmung dieser Retest-Reliabilität erfolgte mittels Berechnung des Pearson-Korrelationskoeffizienten zwischen den Messwerten aus dem Jahr 2012 und jenen aus dem Jahr 2015. Die Korrelation beträgt $r = 0,66$ und ist auf dem Ein-Prozent-Niveau statistisch signifikant. Die Retest-Reliabilitäten der einzelnen Subskalen sind in Tab. 2.2 aufgelistet. Insgesamt ist die Retest-Reliabilität des IBES auf einem befriedigenden Niveau, d. h., die Messwerte sind zeitlich relativ stabil.

Des Weiteren ist für die Qualität eines Testverfahrens die Vorhersagegüte bzw. prognostische Validität bedeutsam. Mit den Messdaten des IBES aus dem Jahr 2012 sollten die kontraproduktiven Handlungen der Studierenden während des Studiums in den Jahren 2012 bis 2015 vorhergesagt werden. Das Ausmaß der kontraproduktiven oder organisationsschädlichen Handlungen wurde im Jahr 2015 erfasst. Hierzu wurde den Studierenden eine Liste mit 68 verschiedenen Handlungen vorgegeben. Diese Liste wurde mithilfe von Fachkräften aus der FHöV und den Ausbildungsbehörden erstellt. Beispiele für kontraproduktive Handlungen sind:

Tab. 2.2 Retest-Reliabilitäten der Skalen des Inventars berufsbezogener Einstellungen und Selbsteinschätzungen. Berechnet wurde der Korrelationskoeffizient nach Pearson, $N = 92$, alle Zusammenhänge sind auf dem Ein-Prozent-Niveau statistisch signifikant

| Subskalen des IBES | Retest-Reliabilität |
|-------------------------------|---------------------|
| Vertraue | 0,54 |
| Geringe Verbreitung | 0,36 |
| Nichtrationalisierung | 0,65 |
| Verhaltensabsichten | 0,64 |
| Gelassenheit/Selbstwertgefühl | 0,72 |
| Zuverlässigkeit/Voraussicht | 0,47 |
| Vorsicht | 0,65 |
| Zurückhaltung | 0,70 |
| Konfliktmeidung | 0,57 |
| Gesamtwert | 0,66 |

- „Ich habe Kommilitonen, Kollegen, Ausbilder, Dozenten, Bürger oder andere beschimpft oder beleidigt.“
- „Möbel oder anderes Inventar der FHöV, des LAFP oder meiner Praxisbehörde habe ich absichtlich beschädigt oder verschmutzt.“
- „In den Praxisphasen habe ich von einem Bürger Geld, Geschenke oder andere Vorteile angenommen.“
- „Ich habe versucht, Dozenten, Ausbilder oder Verwaltungsmitarbeiter durch Geschenke oder Gefälligkeiten positiv zu stimmen.“
- „Ich habe Büromaterial meines Dienstherrn absichtlich für private Zwecke verwendet.“
- „Dozenten, Ausbildern, Verwaltungsmitarbeitern oder Vorgesetzten habe ich bewusst die Unwahrheit oder Halbwahrheit gesagt, um einen persönlichen Vorteil zu erzielen.“
- „Ich habe Unbeteiligten bewusst von Umständen erzählt, die sie von mir nicht hören sollten (z. B. Einsatzpläne von Radarfallen, Daten aus dem Melderegister oder Kfz-Zulassungsdaten).“
- „Bei einer Reisekosten-, Beihilfe- oder Heilfürsorgeabrechnung habe ich bewusst falsche Angaben gemacht.“
- „Mit meinem Dienstaussweis, mit meiner Uniform oder durch Verweis auf mein Amt habe ich mir einen privaten Vorteil verschafft (z. B. Zugang zu einem Club).“

Der Schweregrad der Handlungen wurde von zwei Dienstrechtsexperten eingestuft. Zwei der 92 befragten Studierenden taten kund, während des Studiums keine der vorgegebenen kontraproduktiven Handlungen begangen zu haben. Im Schnitt wurden 7,9 kontraproduktive Handlungen dokumentiert (Standardabweichung: 5,0), die zumindest einmal ausgeführt worden waren. Zwölf und mehr kontraproduktive Handlungen, die mindestens einmal durchgeführt worden waren gaben knapp 20 % der Studierenden an. Der Maximalwert beträgt 23 Handlungen, die jeweils mindestens einmal ausgeführt worden waren, d. h., dieser Studierende hatte jede dritte der im Fragebogen vorgegebenen Handlungen während des Studiums zumindest einmal ausgeführt.

Zur Quantifizierung der prognostischen Validität des IBES wurde die Korrelation (Pearson) zwischen dem Gesamtwert des IBES aus dem Jahr 2012 und dem Gesamtwert des Fragebogens „kontraproduktive Handlungen“ aus dem Jahr 2015 berechnet. Der Gesamtwert des Fragebogens_2015 wurde derart bestimmt, dass die Antwortkategorien „nie“ mit 1, „einmal“ mit 2, „mehrmals“ mit 3 und „oft“ mit 4 bepunktet und anschließend über die 68 Items die Punkte summiert wurden. Die Korrelation zwischen dem Gesamtwert IBES und dem Gesamtwert Fragebogen beträgt $r = -0,33^3$ und ist auf dem Ein-Prozent-Niveau statistisch signifikant. Werden bei der Berechnung des Gesamtwerts

³Das negative Vorzeichen des Korrelationskoeffizienten bedeutet, dass bei hohen IBES-Werten der Gesamtwert des Fragebogens niedrig ist (also wenig kontraproduktive Handlungen angegeben wurden) und umgekehrt.

Tab. 2.3 Ausmaß kontraproduktiver Handlungen im Jahre 2015 in Abhängigkeit vom Gesamtwert des IBES im Jahr 2012

| | Von den Studierenden wären gerügt worden ... (%-Anteil in der Gruppe) | Von den Studierenden wären disziplinarisch geahndet oder entlassen worden ... (%-Anteil in der Gruppe) |
|--|--|---|
| IBES-Gesamtwert deutlich unter Durchschnitt (n = 14) | 100,0 | 64,3 |
| IBES-Gesamtwert knapp unter Durchschnitt (n = 30) | 70,0 | 46,7 |
| IBES-Gesamtwert knapp über Durchschnitt (n = 33) | 84,9 | 45,5 |
| IBES-Gesamtwert deutlich über Durchschnitt (n = 15) | 53,3 | 46,7 |
| <i>Insgesamt (N = 92)</i> | <i>77,2</i> | <i>48,9</i> |

des Fragebogens die einzelnen kontraproduktiven Handlungen (Items) mit ihrem disziplinarrechtlichen Schweregrad gewichtet, ergibt sich $r = -0,32$.

Zur Veranschaulichung der Vorhersagegüte des IBES wurde Tab. 2.3 erstellt. Hierbei wurden die befragten Studierenden danach aufgeteilt, ob ihr Gesamtwert IBES aus dem Jahr 2012 deutlich unter (Perzentil kleiner gleich 16), knapp unter (Perzentil zwischen 17 und 50), knapp über (Perzentil zwischen 51 und 84) oder deutlich (Perzentil größer 84) über dem Durchschnitt des Gesamtwerts IBES lag. Nachdem man diese vier Gruppen auf Grundlage der Daten des IBES aus dem Jahr 2012 gebildet hatte, wurde ausgezählt, wie viele Studierende der jeweiligen Gruppe im Jahr 2015 im Fragebogen mindestens eine kontraproduktive Handlung angeben haben, die zu einer Rüge oder gar zu einer disziplinarischen Ahndung/Entlassung aus dem Beamtenverhältnis führen würde. Alle Studierenden, die 2012 einen deutlich unterdurchschnittlichen IBES-Gesamtwert erzielt hatten, wären 2015 angesichts ihrer kontraproduktiven Handlungen gerügt worden, 64,3 % disziplinarrechtlich belangt oder gar entlassen. Von den Studierenden, die 2012 überdurchschnittliche Ergebnisse erzielten, würden 2015 hingegen nur 53,3 % gerügt und 46,7 % disziplinarrechtlich geahndet/entlassen.

Die Höhe des Korrelationskoeffizienten zwischen dem Gesamtwert IBES sowie dem Gesamtwert Fragebogen und damit die prognostische Validität des IBES ist für Zwecke der Personalauswahl zufriedenstellend. Bei der Interpretation der Höhe des Korrelationskoeffizienten ist zu berücksichtigen, dass die Studierenden der FHöV bereits einen Ausleseprozess durchlaufen hatten – mit der Überprüfung (wenn auch nicht optimal) der charakterlichen und persönlichen Eignung durch die Einstellungsbehörden. Bewerber mit einer sehr hohen Neigung zu kontraproduktiven Handlungen waren wahrscheinlich

im Auswahlprozess der Einstellungsbehörden nicht erfolgreich und sind deswegen nicht in der Gruppe der Studierenden vertreten. Der IBES konnte insofern bei dieser ausgewählten Stichprobe vermutlich nicht an Personen überprüft werden, die eine sehr hohe Neigung zu kontraproduktiven Handlungen haben. Dieser Umstand führt dazu, dass die prognostische Validität des IBES in dieser Studie tendenziell unterschätzt wird.

2.4 Schlussfolgerungen für die Praxis

Der Integritätstest IBES erweist sich – trotz einiger methodischer Einschränkungen der Studie, wie z. B. Erfassung kontraproduktiver Handlungen mittels Selbstauskunft – in dieser Untersuchung als ein für Zwecke der Personalauswahl zufriedenstellend reliables und valides Instrument zur Vorhersage kontraproduktiver Handlungen (und damit auch der Korruption) bei Studierenden der FHöV NRW. Einer Generalisierung der Erkenntnisse und damit Übertragung auf ähnliche Personengruppen (z. B. Nachwuchskräfte des gehobenen Dienstes in anderen Bundesländern, Auszubildende) steht nichts im Wege. Durch den Einsatz des IBES in Auswahlverfahren kann die Gefahr, Nachwuchskräfte einzustellen, die im Dienst kontraproduktiv – insbesondere korrupt – handeln, deutlich reduziert werden.

Beim Einsatz dieses Integritätstests ist die Gleichbehandlung der Bewerber gewährleistet. Schwerwiegende rechtliche Probleme sind nicht zu erwarten. Die Kosten des Einsatzes dieses Instruments sind im Vergleich zu anderen Auswahlinstrumenten niedrig.

Im Hinblick auf die Kriterien Vorhersagegüte, Objektivität, Rechtssicherheit und Wirtschaftlichkeit ist den Behörden und Kommunalverwaltungen zu empfehlen, diesen Integritätstest – ergänzend zu anderen Auswahlinstrumenten – in Stellenbesetzungsverfahren für Nachwuchskräfte zu verwenden. Die Anwendung dieses oder ähnlicher Tests sollte in Auswahlverfahren eingebunden werden, die den fachlichen Standards der DIN 33430 entsprechen.

Literatur

- Freie und Hansestadt Hamburg. (2013). <http://www.hamburg.de/innenbehoerde/korruption/>. Zugegriffen: 1 Juni 2013.
- Friedrich, C. (2003). Vertrauen als Unternehmensressource. In M. Albert (Hrsg.), *Personal als Resource* (S. 119–139). München: Hampp.
- Frintrup, A., Schuler, H., & Mussel, P. (2004). Gelegenheit macht Diebe? Berufliche Integritätsdiagnostik mit PIA. *Wirtschaftspsychologie aktuell*, 4, 58–61.
- Gourmelon, A. (2016). Können kontraproduktive Handlungen von Beamten mit Integritätstests vorhergesagt werden? Empirische Erkenntnisse zur prognostischen Validität des Integritätstests IBES. *Der Öffentliche Dienst*, 2016(2), 29–42.

- Gourmelon, A., Muthers, C., & Stark, C. (2013). Möglichkeiten und Grenzen von Integritätstests bei der Personalauswahl im öffentlichen Sektor – erste Ergebnisse einer Forschungsstudie. In A. Gourmelon (Hrsg.), *Forschung für die Praxis – neue Erkenntnisse für ein professionelles Personalmanagement* (S. 139–158). Heidelberg: Rehm.
- Hanke, J. (2011). *Die inkrementelle Validität eines Integrity-Tests in Bezug auf Ausbildungserfolg – Kann ein Integrity-Test ein Interview ersetzen?* Dissertation. Bonn: Philosophische Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.
- Hoffmann, J., Mokros, A., & Wilmer, R. (2006). Dimensionen der Devianz. *Polizei & Wissenschaft, 1*, 59–63.
- Joas, H. (2001). *Lehrbuch der Soziologie*. Frankfurt: Campus.
- Koch, N. (2013). *Eignungsdiagnostik in der deutschen Personalauswahl mittels Integrity Tests am Beispiel IBES*. Bachelor-Thesis, Fachhochschule für öffentliche Verwaltung NRW, Studienort Gelsenkirchen.
- Marcus, B. (2000). *Kontraproduktives Verhalten im Betrieb*. Göttingen: Hogrefe.
- Marcus, B. (2006). *Inventar berufsbezogener Einstellungen und Selbsteinschätzungen*. Göttingen: Hogrefe.
- Marcus, B. (2007). Neuere Erkenntnisse zum Inventar berufsbezogener Einstellungen und Selbsteinschätzungen (IBES). *Zeitschrift für Personalpsychologie, 6*(3), 129–132.
- Marcus, B., & Wagner, U. (2007). Combining dispositions and evaluations of vocation and job to account for counterproductive work behavior in adolescent job apprentices. *Journal of Occupational Health Psychology, 2*, 161–176.
- Marcus, B., Funke, W., & Schuler, H. (1997). Integrity Tests als spezielle Gruppe eignungsdiagnostischer Verfahren: Literaturüberblick und metaanalytische Befunde zur Konstruktvalidität. *Zeitschrift für Arbeits- und Organisationspsychologie, 41*, 2–11.
- Marcus, B., Schuler, H., Quell, P., & Hümpfner, G. (2002). Measuring counterproductivity: Development and initial validation of a German self-report questionnaire. *International Journal of Selection and Assessment, 10*, 18–35.
- 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*, 1–34.
- Marcus, B., Nijenhuis, J. te, Cremers, M., & Heijden-Lek, K. van der. (o. J.). *Tests of Integrity, HEXACO personality, and general mental ability, as predictors of integrity ratings in the Royal Dutch Military Police*. Unveröffentlichtes Manuskript, FernUniversität Hagen.
- Michaelis, L. O. (2015). Der Einsatz von Integritätstests im Rahmen von beamtenrechtlichen Stellenbesetzungen. *Der Öffentliche Dienst, 9*, 228–236.
- Mussel, P. (2003). Persönlichkeitsinventar zur Integritätsabschätzung (PIA). In J. Erpenbeck & L. von Rosenstiel (Hrsg.), *Handbuch Kompetenzmessung* (S. 3–18). Stuttgart: Schäffer-Poeschel.
- 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), 31–39.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Metaanalysis of integrity test validities. Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679–703.
- Pollmann, A. (2005). *Integrität. Aufnahme einer sozialphilosophischen Personalie*. Transcript: Bielefeld.
- Reinhold, T. (3. März 2007). Der Korruptionstest. *Frankfurter Allgemeine Zeitung, 53*, C4.
- 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*, 262–274.

- Sorgatz, I. (2012). Korruptionsprävention in öffentlichen Institutionen. In A. Gourmelon (Hrsg.), *PÖS – Personalmanagement im öffentlichen Sektor* (Bd. 4). Heidelberg: Rehm.
- Tenzer, E. (2005). Integrität. Was taugt der „Der Weiße-Weste-Check“? *Wirtschaft und Weiterbildung*, 10, 46–49.
- Van Iddekinge, C. H., Roth, P. I., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97, 499–530.
- Zettler, I., & Solga, M. (2008). Instrumente der Arbeits-und Organisationspsychologie. Das „Inventar berufsbezogener Einstellungen und Selbsteinschätzungen“ IBES von Marcus (2006). *Zeitschrift für Arbeits-und Organisationspsychologie*, 52, 97–105.

Korruptionsprävention

Klassische und ganzheitliche Ansätze

Stark, C. (Hrsg.)

2017, XIV, 161 S. 2 Abb., Softcover

ISBN: 978-3-658-06313-9