

Kersting, M. & Klehe, U.-C. (2018). Personalauswahl und -beurteilung: Eine integrative Perspektive. In: S. Greif. & K.C. Hamborg, (Hrsg.), Enzyklopädie der Psychologie. Themenbereich B, Serie III, Band 3, Methoden der Arbeits-, Organisations- und Wirtschaftspsychologie (S. 81-144). Göttingen: Hogrefe

2. Kapitel

Personalauswahl und -beurteilung: eine integrative Perspektive

Martin Kersting und Ute-Christine Klehe

1 Einführung

Right person. Right place. Right time. Das Herzstück guter Personalarbeit besteht in der Fähigkeit Menschen treffend zu beurteilen und zu erkennen, wann jemand wofür „richtig“ gut passt. Sowohl in der Personalauswahl als auch bei Personalbeurteilungen geht es darum, die Eignung zu beurteilen, einmal die Eignung von Bewerbern, einmal die Eignung von Mitarbeitern. Gemeinsam ist beiden Ansätzen auf den ersten Blick der Anforderungsbezug. Weniger salient ist, dass beide Beurteilungssituationen sinnvollerweise eine Förderorientierung erfordern. Nicht nur Mitarbeiter sollten auf der Basis ihrer individuellen Stärken und Entwicklungsbedarfe zielgerichtet gefördert werden, sondern auch Bewerber (angenommene und abgelehnte Bewerber) sollten – nicht nur aus Gründen der sozialen Akzeptanz – zumindest ein förderorientiertes Feedback erhalten. Für ausgewählte Bewerber sollten die Erkenntnisse der Eignungsdiagnostik Grundlage der individuellen Gestaltung des Onboarding Prozesses sowie der Personalentwicklung sein. Auch hinsichtlich der Methoden gibt es Gemeinsamkeiten in den Bereichen Personalauswahl und Personalbeurteilung, beispielsweise hinsichtlich der Skalierung der Leistungen. Den Themen Personalauswahl und -beurteilung ist schließlich gemeinsam, dass berufliche Leistung multifaktoriell bedingt ist. Dies ist sowohl bei der Prognose der Leistung im Kontext der Personalauswahl als auch bei der Beurteilung der Leistung (die zugleich das Kriterium für die Überprüfung der prognostischen Kraft der Personalauswahl darstellt) zu beachten.

Der Beitrag integriert die Ansätze der Personalauswahl und Personalbeurteilung. Zunächst thematisieren wir die Anforderungsanalyse, die die Grundlage sowohl der Personalauswahl als auch der Leistungsbeurteilung ist. Anschließend thema-

tisieren wir zunächst die Personalauswahl und dann die Leistungsbeurteilung. Innerhalb jeder dieser Abschnitte wiederum konzentrieren wir uns erst auf die Konstrukte, die in dem entsprechenden Themenbereich relevant sind, bevor wir uns den jeweiligen Methoden zuwenden, mit denen man versucht, dieser Konstrukte habhaft zu werden.

2 Arbeits- und Anforderungsanalyse als Grundlage der Personalauswahl und -beurteilung

In der Personalauswahl und Personalbeurteilung geht es darum, die Eignung von Bewerbern oder Mitarbeitern zu beurteilen. Der Begriff Eignung bezeichnet terminologisch eine mindestens zweistellige Relation. Anders als bei einstelligen Relationen (wie z. B. der Eigenschaftsbegriff im Sinne von „Frau Müller ist intelligent“), umfasst eine Eignungsbeurteilung notwendigerweise Annahmen über Personen *und* Anforderungen. Mit dem Begriff „Eignung“ werden im Kontext der Personalpsychologie Merkmale einer Person in Bezug gesetzt zu Anforderungen in Ausbildung und Beruf. In der DIN 33430 (2016, S. 7) wird Eignung definiert als:

„Grad der Ausprägung, in dem eine Person über die Eignungsmerkmale verfügt, die Voraussetzung für die jeweils geforderte berufliche Leistungshöhe sind und Zufriedenheit mit dem zu besetzenden Arbeitsplatz, dem Aufgabenfeld, der Ausbildung bzw. dem Studium oder dem Beruf ermöglichen“.

Dem Begriff der Eignung ist der Begriff der Anforderungen inhärent, daher ist der Eignungsbegriff relational und hat eine Leerstelle. Geeignet ist man immer nur „für“ etwas. Die Grundlage jeder Eignungsbeurteilung besteht darin, diese Leerstelle zu sättigen und die Anforderungen festzulegen, denen eine „geeignete“ Person gerecht werden muss. Daher ist für die Personalauswahl und Personalbeurteilung die Definition von Anforderungen obligatorisch. Die Definition der Eignungsmerkmale ist das Ergebnis einer Anforderungsanalyse (vgl. Abb. 1).

2.1 Anforderungsanalytisches Vorgehen

Die Anforderungsanalyse ist die „Systematische Analyse der Anforderungen und der Motivations-/Demotivationspotenziale der beruflichen Tätigkeiten mit dem Ziel der Ermittlung der Eignungsmerkmale von Personen, die bedeutsam dafür sind, dass die Person die erforderliche Leistung erbringt und mit dem zu beset-

Diese Festlegung der notwendigen Eignungsmerkmale oder Soll-Werte kann beispielsweise in Form von Beurteilungskriterien, Anforderungs- oder Kompetenzprofilen erfolgen. Auch andere Bereiche der Personalarbeit wie z. B. die Personalplanung, das Personalmarketing und die -rekrutierung, die Arbeitsplatzgestaltung, der Arbeitsschutz, die Vergütung oder die Personalentwicklung sind auf valide Informationen über die Anforderungen angewiesen (vgl. Brannick, Levine & Morgeson, 2007 sowie Schuler, 2014, für eine Darstellung verschiedener Verwendungsmöglichkeiten von Anforderungsanalysen). Die Anforderungen an eine Person definieren zugleich die Belastung dieser Person sowie das Risiko (z. B. Unfallrisiko) an diesem Arbeitsplatz, so dass die Anforderungsanalyse nicht nur für die Festlegung der Eignungsmerkmale, sondern auch für den Arbeits- und Gesundheitsschutz relevant ist.

Es gibt ein umfangreiches Repertoire an anforderungsanalytischen Methoden (für einen Überblick vgl. Cascio & Aguinis, 2011 sowie z. B. Schuler, 2014). Schuler (2014) differenziert die anforderungsanalytischen Ansätze einerseits danach, welche Grundlage die Erkenntnisse haben und betrachtet andererseits, welche Ebene einer Anforderung beschrieben wird. Bezüglich der Erkenntnisgrundlage unterscheidet er die erfahrungsgeleitet-intuitive Methode, die personenbezogen-empirische Methode und die arbeitsplatzanalytisch-empirische Methode. Jede Studie zur Kriteriumsvalidität ist, was selten expliziert wird, zugleich ein personenbezogen-empirischer Beitrag zur Anforderungsanalyse. In Bezug auf die Beschreibungsebene differenziert Schuler (ebd.) die aufgabenbezogene, verhaltensbezogene und eigenschaftsbezogene Anforderungsanalyse. Innerhalb jeder Kategorie können wiederum eine Vielzahl von Ansätzen differenziert werden. So lassen sich sowohl standardisierte Fragebogen als auch weitgehend freie Verfahren wie anforderungsanalytische Workshops nutzen, um eine verhaltensbezogene Anforderungsanalyse durchzuführen. Die genutzten Verfahren sind von sehr unterschiedlicher psychometrischer Dignität. Hoher Beliebtheit in der Praxis erfreuen sich verhaltensorientierte Methoden wie die Critical Incident Technique (Flanagan, 1954; vgl. dazu Kapitel 1 des vorliegenden Bands). Critical Incidents sind für die Ausübung einer Berufstätigkeit erfolgskritische Situationen. Im Rahmen der Anforderungsanalyse schildern Personen, die mit den beruflichen Anforderungen besonders vertraut sind, Situationen sowie das konkrete Verhalten der Zielpersonen in dieser Situation. Klassische Fragen dabei sind:

1. Was war die Situation, einschließlich relevanter Hintergrundinformation?
2. Was hat der Mitarbeiter in dieser Situation getan?
3. Was waren die Konsequenzen dieses Verhaltens und warum war dieses Verhalten besonders effektiv oder ineffektiv?

Die Situationen können für die Gestaltung von Personalauswahlverfahren (z. B. situative Fragen im Interview, Rollenspiele oder Aufgaben von Situational Judgement Tests) und die Verhaltensschilderungen für die Konstruktion von Verhaltensankern genutzt werden. Die Verhaltensanker werden häufig als Operationalisierung von abstrakt formulierten Anforderungsdimensionen genutzt, indem z. B. die abstrakte Anforderungsdimension „soziale Kompetenz“ durch die Verhaltensanker „hält Blickkontakt“, „monologisiert nicht“, „knüpft an die Ausführungen der Bezugsperson an“ operationalisiert wird. Häufig werden im Rahmen der Personalauswahl oder Personalbeurteilung die Verhaltensanker auch direkt für die Beurteilung genutzt, indem entweder in einem dichotomen Format (Verhalten gezeigt oder nicht gezeigt) oder mit mehreren Skalenstufen im Likert-Format die Beurteiler einschätzen, ob (und in welchem Ausmaß) das Verhalten gezeigt wurde (verhaltensorientierte Einstufungsskalen, Behaviorale Observation Scale; vgl. Abschnitt 6.1.2.2 in diesem Kapitel).

Anforderungsanalysen sind häufig „Einzelfälle“ in dem Sinne, dass in einer Organisation mit spezifischen anforderungsanalytischen Methoden eine Anforderungsanalyse durchgeführt wird. Die Generalisierbarkeit der Erkenntnisse ist fraglich. Um sicher zu stellen, dass keine relevanten Eignungsmerkmale vergessen werden, kann es sinnvoll sein, sich an etablierten anforderungsanalytischen Systemen zu orientieren. Ein solches System liegt dem sogenannten O*Net des U. S. Department of Labor zugrunde (vgl. <http://www.onetcenter.org/>), einer empirisch fundierten Online-Datenbank von kategorisch und hierarchisch geordneten Eignungsmerkmalen einerseits sowie diesen Eignungsmerkmalen entsprechenden anforderungsanalytischen Beschreibungen von Berufen andererseits. Die Datenbank ermöglicht einen Zugriff auf die Anforderungen sowohl einzelner Berufe als auch so genannter „job zones“. Zur Bildung von „job zones“ wurden verschiedene Berufe hinsichtlich des Anforderungsniveaus (Umfang an Ausbildung, Erfahrung usw.) zusammengefasst. Die Kategorisierung von Tätigkeiten nach deren Komplexität sollte bei der Formulierung von Eignungsmerkmalen berücksichtigt werden. Ein dafür häufig genutztes System stellt die sogenannte ISCO-88-Kodierung dar (Elias & Birch, 1994). Hierbei wird angegeben, welches Bildungsniveau für das Erlernen einer Tätigkeit vorausgesetzt wird. Diese Klassifizierung liegt auch für in Deutschland vorkommende Berufe vor (vgl. Geis, 2006).

2.2 Qualifikationen, Kompetenzen und Potenziale

Die Formulierung von Eignungsmerkmalen zielt u. a. darauf ab, das Verständnis von Leistungen und Leistungsvoraussetzungen innerhalb einer Organisation zu vereinheitlichen. Es ist nicht ohne Ironie, dass ausgerechnet die Methoden

und die Begriffe, mit denen Eignungsmerkmale analysiert und beschrieben werden, überaus uneinheitlich und Gegenstand heftiger fachlicher Kontroversen sind. Umstritten sind insbesondere die Gemeinsamkeiten oder Unterschiedlichkeiten der Begriffe Kompetenz einerseits sowie Eigenschaft andererseits. Hier und im Folgenden sprechen wir von Eigenschaften als „Gegenbegriff“ zum Kompetenzbegriff und meinen damit Wissen, Fähigkeiten, Fertigkeiten, Eigenschaften und Motive. Im internationalen Bereich wird häufig von KSAOs gesprochen, die zum festen Begriff gewordene Abkürzung steht für „knowledge, skills, abilities and other characteristics“.

Kompetenzen werden häufig als spezifische, in bestimmten beruflichen Situationen als erfolgsrelevante Kombinationen von Eigenschaften verstanden (z. B. Champion et al., 2011). Bei der in der Praxis häufig anzutreffenden Entwicklung von Kompetenzmodellen werden, anders als beim eigenschaftsorientierten Anforderungsanalytischen Ansatz, nicht nur die aktuellen Anforderungen an Arbeitsplätze berücksichtigt, sondern auch das Leitbild der Organisation, deren strategische Ausrichtung und die zukünftigen Anforderungen. Diesbezügliche Informationen werden u. a. aus Workshops mit Fokusgruppen (z. B. Vertretern des Top-Managements), Kundenbefragungen, aus einer Analyse des Leitbilds und aus Wettbewerberanalysen abgeleitet. Die Kompetenzen der Mitarbeiter gelten als Voraussetzung dafür, die Organisationsziele zu erreichen. Von daher gibt es einen direkten Zusammenhang zwischen Kompetenzmodellen und Modellen der Arbeitsleistung. Ein Beispiel ist das Modell von Bartram (2005), das wir weiter unten im Abschnitt über Arbeitsleistung (vgl. Abschnitt 5.2.1 sowie Abschnitt 6 in diesem Kapitel) vorstellen. In diesem Modell werden u. a. acht Kompetenzen („great eight“) formuliert, denen Verhaltensweisen zugeordnet sind, die der Erreichung der Arbeitsziele dienen.

Die Forderung nach Organisationsspezifität steht im Gegensatz zur Annahme generalisierbarer und somit organisationsübergreifender Fähigkeiten und Eigenschaften. Kompetenzmodelle resultieren häufig aus einem strategieorientierter top-down Ansatz, während eigenschaftsorientierte Anforderungsprofile häufig einem empirieorientierten bottom-up Ansatz entstammen. Insgesamt besteht das Dilemma, dass „klassische“, generisch und situationsunabhängig formulierte Eigenschaften für die Praxis zu abstrakt und praxisfern sind, während organisationsspezifisch formulierte Kompetenzen ggf. nicht situationsübergreifend funktionieren. Es empfiehlt sich daher, beide Ansätze zu berücksichtigen.

Während der Kompetenzansatz in der (Berufs-)Pädagogik verankert ist, stützt sich die klassische Anforderungsanalyse auf die (Eigenschafts-)Psychologie. McClelland (1973) hat seinen Kompetenzbegriff geradezu als Kontrast zum Begriff der Fähigkeiten entwickelt. Aus dieser Tradition heraus wird häufig eine

klare Abgrenzung von Kompetenzen einerseits und Eigenschaften andererseits gefordert. Auf theoretischer Ebene können Kompetenzen von Eigenschaften getrennt werden und zahlreiche Autoren bestehen auf diese Trennung. Demzufolge beziehen sich Kompetenzen in der Regel auf konkrete Situationen (z. B. Führungssituationen), sind also anders als Eigenschaften weniger generalisierbar und können – anders als stabile Eigenschaften – erlernt werden (Erpenbeck, Rosenstiel, Grote & Sauter, 2017). Gerade der letzte Aspekt trägt wesentlich zur Attraktivität des Kompetenzbegriffs bei, da sich durch diese Annahme die Fördermaßnahmen im Rahmen der Personal- und Führungskräfteentwicklung rechtfertigen lassen. In der Praxis spielt die Feindifferenzierung zwischen Kompetenzen einerseits und Eigenschaften andererseits kaum eine Rolle. Fähigkeitsbegriffe sind zentraler Bestandteil zahlreicher Kompetenzmodelle. Ähnlich verhält es sich mit Eigenschaften; nach einer Studie der Personalberatung Strametz (Littau & Hattendorf, 2013) definierten ca. 45 % der befragten Personalleiter und -mitarbeiter den Begriff Kompetenz als „... feste Persönlichkeitseigenschaft, die positiv wirkt“. Kurz und Bartram (2001) weisen auf Zusammenhänge zwischen Kompetenzen und Persönlichkeitseigenschaften sowie Motiven hin. Persönlichkeitseigenschaften können als Grundlage von Kompetenzen angesehen werden (McCrae & Costa, 1997).

In der DIN 33430 (2016) wird der Oberbegriff „Eignungsmerkmal“ genutzt und in Qualifikationen, Kompetenzen und Potenziale differenziert (vgl. Abb. 2). Kompetenzen sind nach DIN 33430 (ebd., S. 7) „Gelernte, wiederholbare Verhaltensweisen und abrufbare Wissensbestände zur erfolgreichen Bewältigung beruflicher Aufgaben“.

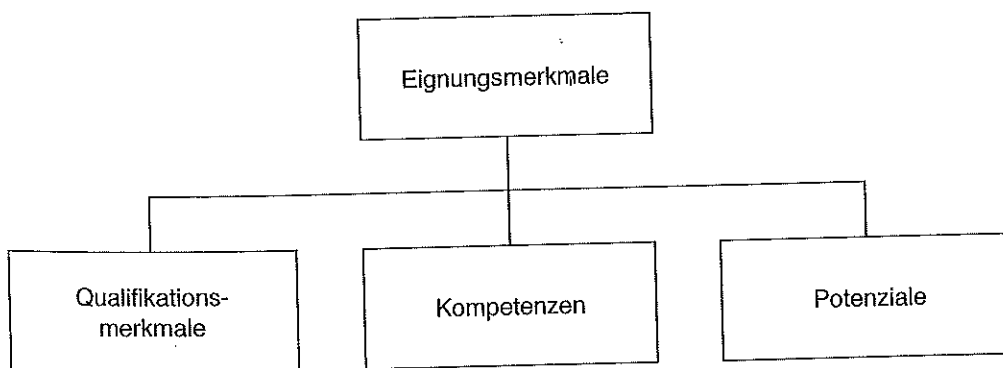


Abbildung 2:
Eignungsmerkmale

Die Erfassung der individuellen Ausprägung von Kompetenzen ist laut DIN 33430 (ebd.) vor allem für die Eignungsbeurteilung von Personen relevant, die die in Frage stehenden beruflichen Aufgaben bereits aktuell ausüben oder ohne

weitere Entwicklungsmaßnahmen übernehmen sollen. Damit ist die Beurteilung von Kompetenzen insbesondere für die Leistungsbeurteilung relevant. Das Potenzial bezeichnet nach DIN 33430 (ebd., S. 8) die „Fähigkeit einer Person, ihr bislang nicht vertraute Aufgaben zu bewältigen und Kompetenzen zu entwickeln“. Anders als Kompetenzen sind Potenziale vor allem für die Eignungsbeurteilung von Personen relevant, die an neue Aufgaben herangeführt werden sollen. Ungeachtet dessen sollten aber natürlich auch bei der Leistungsbeurteilung Potenziale und bei der Eignungsbeurteilung Kompetenzen berücksichtigt werden. Anforderungen an die Qualifikation (z. B. zweites juristisches Staatsexamen) ergeben sich häufig aus formalen Aspekten der in Frage stehenden Position.

2.3 Erfolgskriterien

Die Anforderungsanalyse führt schließlich auch zur Formulierung der Erfolgskriterien. Diese sind im Rahmen der Personalauswahl auch für die Validierung relevant und dienen als Kriterien, die den Prädiktoren (den Auswahlverfahren) gegenübergestellt werden. Im Rahmen der Mitarbeiterbeurteilung sind die Erfolgskriterien Grundlage für die Formulierung der Mitarbeiterziele sowie für die Bemessung der Sollwert- oder Zielerreichung. Häufig werden Leistungsanreize und Vergütungen an die Erfolgskriterien geknüpft.

Es ist üblich, verschiedenen Personalauswahlverfahren, respektive den aufgrund dieser Verfahren getroffenen Entscheidungen, Kriteriumsvaliditäten zuzuschreiben. Die Validität der Prädiktoren variiert aber in Abhängigkeit von dem jeweils gewählten Kriterium. Diese Tatsache ist ebenso trivial, wie weitgehend unbeachtet. Der Großteil der Studien zur Kriteriumsvalidität im Kontext der Personalauswahl basiert auf dem Kriterium der Leistungseinschätzung durch Vorgesetzte, worauf wir später noch eingehen werden. Welche Kriterien man in welcher Kombination als Indikator für den Ausbildungs- und/oder Berufserfolg akzeptiert, sollte aufgrund der Anforderungsanalyse und nicht aufgrund der Verfügbarkeit von Daten entschieden werden.

2.4 Probleme und Herausforderungen der Anforderungsanalyse

Angesichts des gebetsmühlenartig vorgetragenen Mantras der überragenden Bedeutung der Anforderungsanalyse für eine qualitativ hochwertige Personalauswahl und Personalbeurteilung ist es überraschend und enttäuschend, wie selten zu der Frage der Qualität von Anforderungsanalysen geforscht und wie wenig auf die Praxisprobleme der Anforderungsanalyse eingegangen wird.

Auf einer grundsätzlichen Ebene ist das Prinzip der „Passung“ zu hinterfragen, das der Interpretation von anforderungsanalytischen Ergebnissen häufig zugrunde liegt. Die Idee der Passung geht u. a. auf die Annahmen zum „Person-Environment-Fit“ (Kristof, 1996) zurück. Leistung und Zufriedenheit fallen demnach besonders hoch aus, wenn Person und Umwelt zueinander passen. Die Idee überzeugt, viele Aspekte sind aber empirisch bislang ungeklärt. So ist noch nicht hinreichend untersucht, auf welcher „Ebene“ die Person passen soll (Organisation, Vorgesetzter, Kollege, Team, Job) bzw. welche Ebene insgesamt oder in Bezug auf verschiedene Kriterien (z. B. Zufriedenheit, Commitment) besonders wichtig ist (vgl. z. B. die Metaanalyse von Kristof-Brown, Zimmerman und Johnson, 2005). Ebenso muss noch tiefer erforscht werden, welche Form von „Passung“ (z. B. „supplementary fit“ oder „complementary fit“) sich wie auswirkt und ob bestimmte Arten der Passung sich nicht in Form einer zu hohen Homogenität (Schneider, Goldstein und Smith, 1995) und (mangelhaften) „diversity“ (vgl. Kersting & Ott, 2016) negativ auswirken.

Weitere Herausforderungen ergeben sich im anforderungsanalytischen Kontext auf theoretischer Ebene bei der Berücksichtigung der Dynamik. Sowohl die Anforderungen als auch der Mensch verändern sich, wobei sich beide Variablen wechselseitig beeinflussen.

Zu beklagen ist darüber hinaus, dass die Validität der aus der Anforderungsanalyse abgeleiteten Aussagen selten überprüft wird. Diesbezüglich ist zunächst die zweifelhafte Objektivität der Anforderungsanalyse und in Folge dessen die zweifelhafte Validität der aus der Anforderungsanalyse abgeleiteten Interpretation zu nennen. Die Vielzahl der anforderungsanalytischen Methoden sowie der Gestaltungsmöglichkeiten innerhalb einiger Methoden geht mit einer gering ausgeprägten Objektivität einher. Es ist eher unwahrscheinlich, dass zwei verschiedene Anforderungsanalysen zu einer Tätigkeit zum gleichen Ergebnis führen. Das Ergebnis der Anforderungsanalyse variiert beispielsweise in Abhängigkeit von der befragten Gruppe (Kersting & Birk, 2011) sowie der eingesetzten Methoden (Thunsdorff et al., 2015).

Angesichts der hohen Freiheitsgrade im anforderungsanalytischen Vorgehen besteht die Gefahr, dass die Vorstellungen über die Anforderungen vornehmlich Stereotype widerspiegeln. So gehen Eagly und Karau (2002) in ihrer Rollenkongruenztheorie („*role congruity theory*“) beispielsweise davon aus, dass die Unterrepräsentanz von Frauen in Führungspositionen u. a. darauf zurückzuführen ist, dass Führung – und somit die Anforderung an eine Führungskraft – männlich definiert wird („*think-manager-think-male*“, Schein 2001). Für Frauen entsteht ein „Passungsproblem“ („*lack-of-fit*“), da Frauen den Stereotypen zufolge u. a. als „gemeinschaftsorientiert“ und ausgleichend gelten (Williams & Best, 1990).

3 *Ausgewählte Eignungsmerkmale*

Wie oben dargestellt, geht es in der Anforderungsanalyse darum, die Anforderungen einer Tätigkeit zu Merkmalen der Person zu transformieren und diejenigen Eignungsmerkmale (und deren notwendige Ausprägung) zu bestimmen, die notwendig sind, um die Aufgaben unter den gegebenen Bedingungen anforderungsgerecht zu erfüllen und dabei zufrieden zu sein. Dieser Ansatz verkennt keinesfalls, dass über die Eignungsmerkmale hinaus weitere Aspekte, z. B. organisatorische und situative Gegebenheiten, erfolgsrelevant sind. Der Begriff Eignung kann auch für eine drei- oder vierstellige Relation stehen. Im vorliegenden Beitrag konzentrieren wir uns aber auf die Person und ihre Eignungsmerkmale.

Der Möglichkeitsraum der Eignungsmerkmale kann u. a. aus einer (deskriptiven) Analyse der Personalauswahlpraxis oder aus differentiell-psychologischen sowie persönlichkeitspsychologischen Theorien über Personmerkmale abgeleitet werden. Wir betrachten beide Ansätze. Innerhalb der psychologisch fundierten Personmerkmale beschränken wir uns auf vier Bereiche: (kognitive) Fähigkeiten (einschl. Wissen), Motive, Interessen sowie Persönlichkeit im engeren Sinne. Die im Rahmen dieser vier Bereiche formulierten Eignungsmerkmale stellen keinesfalls eine erschöpfende Darstellung aller Eignungsmerkmale dar. Es gibt zahlreiche Eignungsmerkmale auf die wir nicht eingehen, für viele Bereiche ist beispielsweise das motorische Geschick bedeutsam.

Im Folgenden werden für die vier genannten Bereiche (1) „Fähigkeiten“, (2) „Motive“, (3) „Interessen“ sowie (4) „Persönlichkeit im engeren Sinne“ wichtige theoretische Modelle angeführt. Der Rekurs auf die theoretische Basis erlaubt es den Lesern, einzelne diagnostische Verfahren den jeweiligen Konstrukten zuzuordnen und hilft zu entscheiden, ob ein Eignungsmerkmal eher „breit“ oder aber eher „schmal“ (spezifisch) formuliert werden sollte.

Die hier zunächst vorgenommene Ordnung nach Konstrukten steht im Kontrast zu einem Großteil der Literatur zur Personalauswahl, die nach Verfahren geordnet ist. Gängige Gliederungsgesichtspunkte sind bei diesem Zugang Verfahrenskategorien wie „Interview“ und „Assessment Center“ (AC). Die Ordnung nach Verfahren vernachlässigt die Tatsache, dass mit einem Verfahren (wie z. B. einem Interview) unterschiedliche Konstrukte erfasst werden können. Zu Beginn der Gestaltung der Personalauswahl steht die Frage, welche Eignungsmerkmale (also Konstrukte) erfasst werden sollen. Die Verfahrensauswahl ist der zweite Schritt, nachdem die Anforderungsmerkmale in Personmerkmale (Konstrukte) transformiert wurden. Dieser Perspektive entsprechend erfolgt die Darstellung der Konstrukte relativ ausführlich, während die sich anschließende Übersicht über gängige Verfahren knapp ausfällt.

Mit dem vorliegenden Beitrag informieren wir über die Validität von Eignungsdiagnosen. Der Begriff Validität bezieht sich, wie Cronbach (1971, S. 447) bereits anmerkte, immer auf die Interpretation von Daten, die mit einem Verfahren gewonnen wurden, niemals auf ein Verfahren. Ungeachtet dessen wird häufig davon gesprochen, ein Verfahren sei „valide“ oder einem Verfahren würde es an Validität mangeln. Diese Formulierung ist grob vereinfachend und befördert Verständnisprobleme, geht es doch nicht um die Validität eines Verfahrens, sondern um die Validität der Interpretation der Daten, die mit einem Verfahren gewonnen werden. Diese Erkenntnis war einige Jahrzehnte in Vergessenheit geraten, setzt sich nun aber wieder durch (Kane, 2013) und wurde erfreulicherweise mit Nachdruck in die aktuelle Fassung der „Standards for Educational and Psychological Testing“ aufgenommen (American Educational Research Association, Psychological Association & National Council on Measurement in Education, 2014).

Interpretiert werden Messwerte. Die Interpretation eines Messwertes ist eine unauflösbare Einheit aus einem Konstrukt einerseits und einer Methode andererseits. Berichte über die Validität von Interpretationen müssen daher über Konstrukt und Methode informieren. Leider ist die Literatur zur Validität im Kontext der Personalauswahl nicht immer hinreichend informativ. So finden sich zahlreiche Berichte über die Validität der aus ACn oder Interviews abgeleiteten Interpretationen, ohne dass ausreichend über das jeweils erfasste Konstrukt und dessen Anforderungsrelevanz berichtet wird. Ein vollständiges Bild über die Validität ergibt sich nur, wenn man Konstrukte und Verfahren simultan betrachtet. Entsprechend dieser Befundlage berichten wir im vorliegenden Buchbeitrag sowohl in den nach Konstrukten geordneten Abschnitten als auch in einzelnen verfahrensbezogenen Abschnitten Befunde zur Validität.

3.1 Eignungsmerkmale, die in der Praxis häufig für relevant erachtet werden

Um einen Einblick zu erhalten, welche Eignungsmerkmale im Rahmen der Personalauswahl in der Praxis häufig geprüft werden, kann man sich anschauen, welche Eignungsmerkmale beispielsweise in einem AC beobachtet werden. Dies wurde in verschiedenen Studien ausgewertet. Da nur anforderungsrelevante Merkmale beobachtet werden sollen, geben diese Studien einen Überblick über in der Praxis genutzte Eignungsmerkmale. Arthur, Day, McNelly und Edens (2003) haben auf der Basis von 34 Studien 168 verschiedene AC Dimensionen identifiziert und zu letztendlich sechs generischen Kompetenzen zusammengefasst: communication (Kommunikationsfähigkeit), consideration and awareness of others (Vergegenwärtigung/Berücksichtigung anderer), drive (Antrieb), influencing other (Beeinflussung anderer), organizing and planning (Organisation und Planen) und problem solving (Problemlösen).

In einer vergleichbaren Arbeit mit dem Fokus auf Development Center und Personalentwicklung reduzierten Gibbons, Rupp, Snyder, Holub & Woo (2006) über 1 000 Kompetenzen auf 16 Dimensionen, die sie wiederum zu vier Clustern zusammenfassten. Drei der vier Cluster weisen hohe Überlappungen zu den von Arthur und Kollegen (2003) formulierten Kompetenzen Kommunikationsfähigkeit, Vergegenwärtigung/Berücksichtigung anderer und Problemlösen auf. Das vierte, „approach to work“ genannte, Kompetenzcluster fasst die bei Arthur und Kollegen (2003) separierten Dimensionen Antrieb, Organisation und Planen sowie Stresstoleranz zusammen.

Mit drei Kompetenz-Dimensionen zur Beschreibung von AC Leistungen kommen Kolk, Born und van der Flier (2004) aus, sie nehmen eine Unterteilung in „feeling“ (zwischenmenschliche Fähigkeiten), „thinking“ (intellektuelle Fähigkeiten) und „power“ (Durchsetzungskraft) vor.

3.2 Theoretisch fundierte Eignungsmerkmale

Auch wenn das Vorgehen bei der Erarbeitung der relevanten Eignungsmerkmale (Kompetenzmodellentwicklung versus Anforderungsanalyse) unterschiedlich ist, die resultierenden Merkmale sind offensichtlich ähnlich. Diese Merkmale lassen sich problemlos in psychologische Theorien einordnen. Eignungsmerkmale wie Kommunikationsfähigkeit, Vergegenwärtigung/Berücksichtigung anderer und Problemlösefähigkeiten/analytische Fähigkeiten sind genuiner Bestandteil gängiger differentiell-psychologischer oder persönlichkeitspsychologischer Theorien über Personmerkmale, werden dort in der Regel aber weniger „operativ“ gefasst. So ist die Kommunikation eher dem breiteren Faktor „Extraversion“ zugeordnet und das Merkmal „Berücksichtigung anderer“ der Verträglichkeit. Dennoch zeigt sich, dass die unterschiedlichen Traditionen der Kompetenzmodelle einerseits und der Anforderungsprofile andererseits trotz aller theoretischen Unterscheidungen auf der Ebene der Eignungsmerkmale praktisch konvergieren.

3.2.1 Kognitive Fähigkeiten

Auf der Basis von Tests zu kognitiven Fähigkeiten getroffene Entscheidungen erzielen die höchste Treffsicherheit bei der Vorhersage bildungs- und berufsrelevanter Erfolgskriterien. Dies gilt für den Erfolg in der Schule, in der beruflichen Ausbildung, im Studium und im Beruf. Schmidt und Hunter (1998) bestimmten die mittlere korrigierte Validität von Intelligenztests in Bezug auf den allgemeinen Berufserfolg mit $r = .51$; Salgado, Anderson, Moscoso, Bertua und De Fruyt (2003a) schätzen diesen Wert auf $r = .62$, wenden allerdings diskussionsbedürftige Korrekturmethode an. Ein Bericht über metaanalytisch ermit-

telte Kriteriumsvaliditäten einzelner Intelligenzdimensionen (z. B. sprachgebundenes Denken, rechnerisches Denken) geben Goertz, Hülshager und Maier (2014); Lang, Kersting, Hülshager und Lang (2010) sowie Salgado und Kollegen (2003a).

Metaanalysen zu deutschen Studien zur Kriteriumsvalidität von Intelligenztests wurden von Hülshager, Maier und Stumpp (2007) sowie von Kramer (2009) publiziert. Die zuerst genannten Autoren kommen auf einen Vorhersagewert von .47 für Ausbildungserfolg und .53 für Berufserfolg, Kramer (ebd.) berichtet eine operationale Validität in Bezug auf berufsbezogene Lernleistungen in Höhe von .62.

Dabei erlauben Intelligenztests generell – also über alle Berufe hinweg – valide Schlussfolgerungen (Schmidt et al., 1993; Salgado et al., 2003b). Empirisch lässt sich zeigen, dass Intelligenz umso bedeutsamer ist, umso komplexer die beruflichen Herausforderungen sind (Lang et al., 2010; Salgado et al., 2003b). Intelligenztests erlauben langfristige Prognosen (Judge, Higgins, Thoresen & Barrick, 1999).

Als Fazit jahrzehntelanger Forschung formulieren Schmidt und Hunter (2004, S. 162) in Bezug auf die Eignungsdiagnostik in ihrem Überblicksartikel zur Bedeutung der Intelligenz (general mental ability, GMA abgekürzt) „the research evidence that GMA predicts both occupational level attained and performance within one's chosen occupation and does so better than any other ability, trait, or disposition and better than job experience. The sizes of these relationships with GMA are also larger than most found in psychological research.“

Es besteht ein großes Angebot an Intelligenz- und Fähigkeitstests für die Personalauswahl. Um die Tests einordnen und um eine Testauswahl treffen zu können, empfiehlt es sich, sich an einem etablierten integrativen Modell intellektueller Fähigkeiten zu orientieren. Es gibt zahlreiche Modelle, die dafür in Frage kommen. Sehr gut geeignet ist das Cattell-Horn-Carroll-Modell (CHC-Modell, McGrew, 2009; Schneider & McGrew, 2012), da es in hoher Ausdifferenzierung relativ viele Fähigkeiten (inkl. Wissen als „job knowledge“) berücksichtigt, hierarchisch gegliedert und gut beforscht ist.

Eine andere Systematik eröffnet das Berliner Intelligenzstrukturmodell (BIS, Jäger, 1984). Für die Umsetzung der anforderungsanalytischen Ergebnisse in Personmerkmale und für die Zuordnung von diagnostischen Verfahren zu diesen Personmerkmalen – und somit für eine konstruktgetreue Interpretation – sind theoretische Rahmenmodelle unverzichtbar. Eine Zuordnung von Intelligenztests zu Intelligenztheorien liefert die hervorragende Überblicksarbeit von Süß und Beauducel (2011).

3.2.2 *Motive*

Von großer Bedeutung für die Arbeitswelt ist die Motivation der Menschen. Die Motivation ist die Menge aller Motive, wobei die Wechselwirkung zwischen motiviertem Subjekt und motivierender Situation (Graumann, 1969) zu beachten ist. Motive tragen wesentlich dazu bei, zu begründen, „warum“ sich ein Mensch auf eine bestimmte Art und Weise verhält (McClelland, 1985). Persönlichkeitseigenschaften im Sinne des Fünf-Faktoren-Modells (vgl. Abschnitt 3.2.4 in diesem Kapitel) leisten hingegen eher einen Beitrag zur Erklärung, „wie“ ein Mensch seine Ziele zu erreichen sucht. Die Motive werden als Energie konzeptualisiert, während die Persönlichkeitseigenschaften im Sinne des Fünf-Faktoren-Modells wesentlich darüber entscheiden, welche Verhaltensweisen energetisiert werden (Brunstein, 2003; Brunstein und Maier, 2005). Über den Zusammenhang zwischen Motivation und den fünf Faktoren der Persönlichkeit berichten anhand metaanalytischer Daten Judge und Ilies (2002).

Von den zahlreichen Kategorisierungssystemen für Motive soll hier die etablierte Einteilung in das (1) Machtmotiv, (2) Leistungsmotiv und (3) Anschlussmotiv (McClelland, 1985) herausgegriffen werden. Nach dieser Theorie lassen sich bei jedem der drei genannten Motive eine (a) annähernde und (b) eine meidende Tendenz unterscheiden. Entsprechend unterscheidet man die Hoffnung auf (1a) Macht, (2a) Erfolg und (3a) Anschluss einerseits sowie die davon weitgehend unabhängige Furcht vor (1b) Machtverlust, (2b) Misserfolg und (3b) Zurückweisung andererseits. Schließlich werden Motive danach unterschieden, ob sie implizit oder explizit sind – dieser Punkt ist von zentraler Bedeutung für die Personalauswahl und wird weiter unten ausführlicher thematisiert. Zahlreiche Forscher gehen davon aus, dass mit den impliziten und expliziten Motiven zwei unabhängige Motivsysteme existieren (vgl. z. B. Back, Schmukle & Egloff, 2009; McClelland, Koestner und Weinberger, 1989). Während explizite Motive als Ursache für explizit von „außen“ gefordertes (respondentes) Verhalten gelten, sind implizite Motive für spontan auftretendes (operantes) Verhalten verantwortlich.

Unter dem Motivationsbegriff in einem weiteren Sinne lassen sich auch Interessen, Einstellungen, Werte, Selbstbilder und gelernte Normen sowie Rollenbilder subsumieren, die eher mit den expliziten als mit den impliziten Motiven einer Person zusammenhängen dürften.

Es gibt nur wenige empirische Studien zur Frage, inwieweit eine Diagnostik der Motivation im Rahmen der Personalauswahl valide Schlussfolgerungen erlaubt. Ein Großteil der Studien, deren Ergebnisse sich auf die Personalauswahl übertragen lassen, ist veraltet. Die „klassische“ Studie von McClelland und Kollegen zur langfristigen Prognose des Erfolgs von Managern in Abhängigkeit von deren

mit projektiven Verfahren gemessenen impliziten Motiven stammt aus den 1980er Jahren (z. B. McClelland & Boyatzis, 1982). Die Studie gilt sowohl als Beleg für die Annahme, dass mit Verfahren zur impliziten Motivdiagnostik längerfristige Prognosen getroffen werden können als mit Verfahren zur expliziten Motivdiagnostik als auch für die Annahme, dass erfolgreiche Führungskräfte eine bestimmte Konstellation von impliziten Motiven, nämlich das „Leadership Motive Pattern“ aufweisen. Damit ist die Kombination eines hoch ausgeprägten Macht-, eines mittel ausgeprägten Leistungs- und eines gering ausgeprägten Anschlussmotivs gemeint (vgl. auch Yukl, 2012).

Spangler (1992) konnte metaanalytisch zeigen, dass das mit dem TAT erfasste Leistungsmotiv substanziell mit verschiedenen Außenkriterien (wie z. B. Ausdauer) korrelierte und diese Korrelation teilweise höher ausfiel als die entsprechenden Korrelationen für die Interpretation der mit herkömmlichen Fragebogen erfassten expliziten Leistungsmotive. Zu ähnlichen Ergebnissen kommen McClelland und Franz (1992). An aktuellen Studien kann die Metaanalyse von Collins, Hanges und Locke (2004) aufgeführt werden. Die Autoren zeigten metaanalytisch einen Zusammenhang zwischen Leistungsmotivation einerseits und der Entscheidung Entrepreneur zu werden sowie mit Indikatoren der Leistung als Entrepreneur andererseits. Dabei konnte das Leistungskriterium überraschenderweise besser vorhergesagt werden als die Karriereentscheidung (.46 zu .21). Der Befund einer höheren Aussagekraft der impliziten Motive, der sich in der Studie von Spangler (1992) gezeigt hatte, konnte in der Metaanalyse von Collins und Kollegen (2004) nicht repliziert werden.

Trotz der heterogenen Befundlage ist davon auszugehen, dass berufsrelevantes Verhalten wesentlich von impliziten Motiven bestimmt wird. Eine Motivationsdiagnostik im Rahmen der Personalauswahl müsste daher nicht nur explizite, sondern auch implizite Motive erfassen. Implizite Motive lassen sich aber nicht durch herkömmliche diagnostische Methoden wie z. B. Fragebogen oder Interviews erfassen. Zur Diagnostik impliziter Motive bedarf es indirekter Verfahren (für einen Überblick vgl. Schmukle & Egloff, 2011). Trotz berechtigter Kritik an deren Objektivität und Validität können projektive Verfahren (z. B. der thematische Apperzeptionstest, TAT), semi-projektive Verfahren (z. B. das Multi-Motiv-Gitter, MMG), objektive Tests sensu Cattell (z. B. der Objektive Leistungsmotivationstest, OLMT) und implizite Assoziationstests (IAT) als geeignete Instrumente der Motivdiagnostik bezeichnet werden. Dem Einsatz dieser Verfahren in der Personalauswahl stehen aber rechtliche Gründe entgegen. Darüber hinaus dürfte die Akzeptanz solcher Verfahren in der Personalauswahl u. a. aufgrund der mangelhaften Transparenz gering ausfallen. Im Rahmen der Personalauswahl begnügt man sich daher häufig damit, in standardisierter (z. B. mit Hilfe eines Motivationsfragebogens wie dem Leistungsmotivationsinventar, LMI) oder halb-standardisierter (z. B. in einem Interview) Form nach der Moti-

vation zu fragen. Damit erfasst man aber nur explizite Motive, Motive also, die sich eine Person bewusst zuschreibt. Die Ergebnisse der Messungen dieser expliziten Motive sind weitgehend unabhängig von den Ergebnissen in Tests zur Erfassung impliziter Motive.

Noch ein weiterer Gesichtspunkt erschwert die Diagnose von Motiven: Der Zusammenhang zwischen Motiven und Leistung wird durch weitere Variablen wie z. B. der Aufgabenschwierigkeit und den jeweiligen Anreizen moderiert, dieser Aspekt muss bei der Interpretation von Motivationsindikatoren berücksichtigt werden.

3.2.3 Berufliche Interessen

Berufliche Interessen wurden bislang eher Bedeutung für die Studienfachorientierung und die Berufswahlberatung, weniger aber für die Personalauswahl zugesprochen. Dies ändert sich, seitdem die klassische Personalauswahl immer häufiger um vorgelagerte „Self-Assessments“ ergänzt wird, bei denen Personen vor der eigentlichen Bewerbung (und zumeist anonym) in einem – in der Regel online ohne Aufsicht ablaufenden – Test ausprobieren können, ob ihr Eignungsprofil zu dem Anforderungsprofil passt (vgl. Ott, Ulfert und Kersting, 2017). Dies soll die Selbstreflexion über die Passung von persönlicher Eignung und Anforderungen anregen und eine stimmige Selbstselektion hervorrufen. In diesem Kontext kommt den Interessen eine hohe Bedeutung zu.

Nach der Kongruenztheorie von Holland (1997) geht eine hohe Übereinstimmung zwischen den stabilen Interessen einer Person einerseits und der „Umwelt“ (Ausbildungen, Berufe und Arbeitsumgebungen) andererseits mit erhöhter Zufriedenheit einher. Das Modell von Holland unterscheidet die folgenden sechs Interessentypen: R (realistic): Praktisch-technische Orientierung; I (investigative): Intellektuell-forschende Orientierung; A (artistic): Künstlerisch-sprachliche Orientierung; S (social): Soziale Orientierung; E (enterprising): Unternehmerische Orientierung und C (conventional): Konventionelle Orientierung. Sowohl die Person als auch die Umwelt/die Arbeitswelt kann nach diesen Kategorien beschrieben werden.

Die empirische Forschung im Bereich beruflicher Interessen beschränkt sich im Wesentlichen auf die mit Fragebogen erfassten Interessen. Das erleichtert die Interpretation der Daten, die einheitlich einer bestimmten Konstrukt-Methoden Kombination entsprechen. Interessen manifestieren sich aber auch in Handlungen, z. B. dem Aufsuchen bestimmter Umwelten. Entsprechende Informationen können in biografischen Fragebogen oder Interviews berücksichtigt werden, dazu wird aber kaum geforscht.

Die metaanalytischen Befunde von Van Iddekinge, Roth, Putka und Lanivich (2011) stellen die Praxis, den Einsatz von Interessen-Fragebogen auf die Berufsberatung zu begrenzen, in Frage. Die Autoren konnten aufgrund von metaanalytisch über 74 Studien hinweg ermittelten Daten zeigen, dass Interessen verschiedene Kriterien (Arbeitsleistung (.14), Leistung in Aus- und Weiterbildung (.26), Kündigungsabsicht (-.19) sowie Kündigung (-.15)) vorhersagen, so dass Interessensfragebogen prinzipiell auch im Kontext der Personalauswahl genutzt werden können.

Indirekt spielen Interessen auch bei einer auf Intelligenz und Persönlichkeit ausgerichteten Diagnostik eine Rolle, da Interessen mit Persönlichkeitsmerkmalen (vgl. Barrick, Mount & Gupta, 2003) sowie mit intellektuellen Fähigkeiten (vgl. Pässler, Beinicke & Hell, 2015) zusammenhängen.

Sowohl bei der Beratung als auch bei der Personalauswahl auf der Basis von Berufsinteressen muss allerdings berücksichtigt werden, dass Interessen stark geschlechtsstereotyp geprägt sind. Su, Rounds und Armstrong (2009) konnten in einer empirisch überaus fundierten Metaanalyse deutliche geschlechtsspezifische Effekte aufzeigen, demzufolge sich (junge) Frauen eher für „Menschen“ und (junge) Männern eher für „Gegenstände“ interessieren. Eine Auswahl der Berufe oder Bewerber nach Interessen läuft Gefahr, geschlechtsstereotype Berufsvorstellungen zu zementieren. Demgegenüber fällt die Berufswahl weniger geschlechtsstereotyp aus, wenn Menschen ihre Wahl nach ihren Fähigkeiten orientieren. Eine Studie von Zimmer und Sander (2014) zur Wirkung von Self-Assessments zeigt, dass (jungen) Frauen auf der Basis von Interessenstests nur in knapp 6 % der Fälle ein Beruf im technischen Bereich empfohlen wird, gegenüber einer Empfehlungsrate von ca. 20 % bei einer fähigkeitsbasierten Beratung.

3.2.4 Persönlichkeitsmerkmale im engeren Sinne

Selbstverständlich gehören die bislang thematisierten Merkmale der intellektuellen Fähigkeiten, Motive und Interessen zur Persönlichkeit. Dennoch ist es üblich, die drei genannten Konstrukte separat von Persönlichkeitsmodellen im engeren Sinne zu thematisieren, was natürlich nicht ausschließt, über die Zusammenhänge der Konstruktbereiche nachzudenken (vgl. z. B. Ackerman & Heggstad, 1997).

Als Rahmenmodell für die Persönlichkeit im engeren Sinne hat sich, wie oben bereits angedeutet, das Fünf-Faktoren-Modell der Persönlichkeit (FFM) etabliert. Das Verhalten von Menschen lässt sich demzufolge als Indikator für Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit interpretieren, umgekehrt formuliert erhöhen bestimmte Persönlichkeitsausprägungen die Auftretenswahrscheinlichkeit für bestimmte Verhaltensweisen.

Erst durch die theoriegeleitete Integration der Befunde im Sinne des FFM war es möglich, mit Metaanalysen in der Vielzahl von Einzelstudien ein weitgehend konsistentes Bild der Kriteriumsvalidität von persönlichkeitsorientierten Eigenschaftsaussagen aufzuzeigen. Die empirischen Ergebnisse lassen sich gut interpretieren, da sich die berichteten Kennwerte auf eine einheitliche Kombination von einem Messinstrument (Fragebogen) und einem Konstrukt beziehen. Auf der Basis von fünfzehn Metaanalysen bestimmten Barrick, Mount und Judge (2001) in einer „second order“-Metaanalyse den Zusammenhang zwischen verschiedenen Kriterien des Berufserfolges (z. B. Vorgesetztenbeurteilung und Trainingsleistung, differenziert nach verschiedenen Berufsgruppen) und den fünf Persönlichkeitsdimensionen im Sinne des FFM. Die Dimension Gewissenhaftigkeit erwies sich für alle Kriterien und für alle Berufsgruppen als bedeutsam. Auf der Konstruktebene lag der für die Population geschätzte Zusammenhang von Gewissenhaftigkeit mit („overall“) Berufserfolg bei $r = .27$ (berichtet werden hier und im Folgenden vollständig korrigierte Zusammenhangsmaße). Auch der Neurotizismus war zu $r = -.13$ mit dem allgemeinen Berufserfolg (negativ) korreliert, hier zeigten sich Unterschiede in der Bedeutsamkeit der Dimension für einzelne Berufsgruppen. Die Kriteriumsvalidität der Dimensionen Extraversion, Offenheit für Erfahrungen und Verträglichkeit variierte in Abhängigkeit von den jeweils gewählten Kriterien und Berufsgruppen. In einer anderen Metaanalyse (Judge, Bono, Ilies, und Gerhardt, 2002) erwies sich Extraversion mit $r = .31$ als bester Prädiktor für „Führung“, wobei der Zusammenhang erwartungsgemäß für das Kriterium „leader emergence“ höher ausfiel ($r = .33$) als für „leader effectiveness“ ($r = .24$). Weitere bedeutsame Prädiktoren für „Führung“ waren Gewissenhaftigkeit ($r = .28$) und Neurotizismus ($r = -.24$).

Extraversion ist – ebenso wie Neurotizismus ($r = -.29$) – auch ein bedeutsamer Prädiktor der beruflichen Zufriedenheit ($r = .25$), wie Judge, Heller und Mount (2002) metaanalytisch zeigen konnten. Laut Judge und Kollegen (1999) lässt sich die Arbeitszufriedenheit mit Merkmalen der Persönlichkeit im engeren Sinne besser vorhersagen als mit Intelligenz.

Die Befunde zur Bedeutung von Persönlichkeitsmerkmalen für den Berufserfolg sind vor allem vor dem Hintergrund der relativen Unabhängigkeit der fünf Persönlichkeitsfaktoren einerseits und der Ausprägung der kognitiven Fähigkeiten andererseits interessant. Dadurch ergibt sich ein Potenzial für inkrementelle Validität. Tatsächlich sprechen Schmidt und Hunter (1998) den Interpretationen von Fragebogen zur Erfassung der Gewissenhaftigkeit (sowie zur Integrität) inkrementelle Validität gegenüber rein intelligenztestbasierten Interpretationen zu.

Vergleichbar der Diskussion um die spezifische Fähigkeitstheorie im Kontext der Intelligenzforschung wird auch im Bereich der Persönlichkeit die Frage nach

dem angemessenen Spezifitäts- bzw. Generalitätsniveau der Messung von Persönlichkeitsmerkmalen diskutiert. Während z. B. Ones und Viswesvaran (1996) die Nutzung eher globaler, abstrakter Maße favorisieren, sprechen sich z. B. Paunonen und Ashton (2001) explizit für eine spezifische Beschreibungsebene aus. Bezogen auf das FFM würde dies für eine Interpretation auf der Facettenebene („little thirty“) sprechen. Statt der breiten Faktoren wie Gewissenhaftigkeit sollte man also spezifische Facetten wie Leistungsstreben und Selbstdisziplin interpretieren. Insgesamt erscheint die Frage nach dem angemessenen Generalitätsniveau auch in diesem Bereich noch offen. Befunden zur überlegenen Validität der Interpretation auf Facettenebene (z. B. Dudley, Orvis, Lebiecki & Cortina, 2006; Paunonen & Ashton, 2001 sowie – am Beispiel von Verkäufern/Vertrieblern – Vinchur, Schippmann, Switzer & Roth, 1998) stehen Daten gegenüber, die für „compound traits“ hohe Validitäten nachweisen. „Compound traits“ setzen sich aus mehreren Dimensionen des FFM zusammen und sind somit „breit“ konzeptualisiert. Entsprechende Messungen werden häufig im Rahmen sogenannter „Criterion-Focused Occupational Personality Scales“ (COPS, Ones & Viswesvaran, 2001a) realisiert. Ein prominentes Beispiel sind Integrity Tests. Interpretationen der Daten von Integrity Tests weisen in einigen Studien (Marcus, Höft & Riediger, 2006; Ones & Viswesvaran, 2001b) eine hohe Kriteriumsvalidität im Allgemeinen und eine inkrementelle Validität gegenüber Intelligenztests im Besonderen auf (Schmidt & Hunter, 1998). In der Metaanalyse von Van Iddekinge, Roth, Raymark und Odle-Dusseau (2012) zeigten sich für Eignungsaussagen auf der Basis von Integritätstests allerdings nur geringe Kriteriumsvaliditäten ($r = .12$ für das Kriterium „job performance“).

Mit Wittmann (1988) kann auch für diesen Bereich davon ausgegangen werden, dass hohe Kriteriumsvaliditäten dem Symmetrieprinzip zufolge dann erzielt werden können, wenn die Spezifität des Prädiktor-Konstrukts der Spezifität des Kriterium-Konstrukts entspricht. Maximale Validität wäre demnach nur dann zu erzielen, wenn Prädiktor und Kriterium das gleiche Generalitätsniveau aufweisen. In diese Richtung gehen die ermutigenden Ergebnisse, die Hogan und Holland (2003) für Eignungsaussagen auf der Basis spezifischer Persönlichkeitsfacetten und in Bezug auf spezifische Berufserfolgskriterien aufzeigen.

Aktuell zeichnet sich ab, dass das FFM als theoretisches Leitmodell durch das HEXACO-Modell (z. B. Ashton & Lee, 2007; Ashton, Lee, Perugini et al., 2004) abgelöst wird.

Erste Studien erlauben einen vorsichtigen Optimismus, dass sich mit Interpretationen auf der Basis des HEXACO-Modells bessere Kriteriumsvaliditäten als mit Interpretationen auf der Basis des FFM ergeben (Ashton, Lee, Perugini et al., 2004; Marcus, Lee & Ashton, 2007; Johnson, Rowatt & Petrini, 2011; Zettler & Hilbig, 2010).

3.2.5 Weitere Eignungsmerkmale

Weitere Eignungsmerkmale ergeben sich u. a. aus Ansätzen der Erweiterung des traditionellen Intelligenzbegriffs.

Im Kontext der beruflichen Eignung ist hier zunächst die praktische Intelligenz zu nennen, die in zwei verschiedenen Bedeutungen relevant ist. In der ursprünglichen Bedeutung geht es bei der praktischen Intelligenz um die praktisch-mechanisch-räumlich-physikalischen Fähigkeiten im Sinne des entsprechenden Hauptgruppenfaktors (k:m; practical-mechanical-spatial-physical) im hierarchischen Intelligenzmodell Modell von Vernon (1959). Entsprechende Fähigkeiten sind in zahlreichen handwerklichen Berufen gefragt und können sowohl mit Papier-Bleistift-Tests (respektive Computer-Tests) wie dem Mechanisch Technischen Verständnistest (MTVT, Lienert, 1958) als auch mit apparativen Verfahren/Arbeitsproben wie der Drahtbiegeprobe (Lienert, 1967) erfasst werden. Die Kriteriumsvalidität derartiger Testverfahren ist insgesamt – sofern die diagnostizierten Merkmale anforderungsrelevant sind – gegeben (vgl. z. B. Muchinsky, 1993), allerdings zeigen sich hinsichtlich der geschlechtsspezifischen Ausprägungen deutliche Leistungsvorteile für die Männer.

Etwas gänzlich anderes ist mit der praktischen Intelligenz im Sinne von Sternberg und Wagner (1986) gemeint. Die Autoren meinen damit ganz allgemein das „gewusst wie“ einer alltäglichen Problemlösung. Zur Problemlösung braucht es demnach „stilles Wissen“ (tacit knowledge) über das (1) Selbstmanagement (2) das Management anderer sowie (3) das Management von Aufgaben. Zur Erfassung des stillen Wissens im Rahmen der Personalauswahl wurden Situational Judgement Tests (SJTs) entwickelt. Es kann allerdings bezweifelt werden, ob mit SJTs wirklich stilles Wissen erfasst wird. Eine alternative Erklärung geht davon aus, dass mit SJTs explizites Wissen über sozial erwünschtes Verhalten am Arbeitsplatz erfasst wird. Dies legen Studien nahe, die zeigen, dass die SJTs auch ohne die Situationsdarstellung gelöst werden können (Krumm et al., 2015).

Bei einem weiteren Intelligenzbegriff, der sozialen Intelligenz, handelt es sich streng genommen nicht um eine Erweiterung, sondern um einen Ursprung des Intelligenzbegriffs, der auf Thorndike zurückgeht (1920, zitiert nach Weis, Seidel & Süß, 2006) und sowohl Kognitionen als auch Verhalten umfasst. Eine aktuellere und gegenüber der akademischen Intelligenz vom Anspruch her eigenständigere Variante der sozialen Intelligenz ist die emotionale Intelligenz (Mayer & Salovey, 1997). Die Autoren unterscheiden vier Komponenten emotionaler Intelligenz: Die Emotionswahrnehmung, die Emotionsnutzung, das Emotionswissen und das Emotionsmanagement. Es gibt sowohl Ansätze, die emotionale Intelligenz mit Leistungstests zu erfassen als auch Ansätze der Operationalisierung über Selbstberichts-Fragebogen. Ein Beispiel für einen Leistungstest zur emotionalen Intelligenz ist der Mayer-Salovey-Caruso Test zur Emotionalen In-

telligenz (MSCEIT), die deutsche Fassung des MSCEIT wurde von Steinmayr, Schütz, Hertel und Schröder-Abé (2011) verfasst. Ein Beispiel für einen Selbstberichts-Fragebogen ist der Fragebogen zur wahrgenommenen emotionalen Intelligenz von Otto, Döring-Seipel, Grebe und Lantermann (2001).

Die Ergebnisse in Leistungstests und in Selbstberichtsverfahren zur emotionalen Intelligenz hängen nur gering miteinander zusammen, wie u. a. Joseph und Newman (2010) zeigen. Die Autoren formulieren in ihrer empirisch gestützten Arbeit ein Kaskadenmodell der emotionalen Intelligenz mit der Emotionswahrnehmung als erster, dem Emotionsverständnis als zweiter und der Emotionssteuerung als dritter Stufe. In der Metaanalyse zeigten sich deutliche Zusammenhänge zwischen der Gewissenhaftigkeit und der Emotionswahrnehmung, der akademischen Intelligenz und dem Emotionsverständnis sowie zwischen der emotionalen Stabilität und der Emotionssteuerung. Die zur Kriteriumsvalidierung herangezogene Berufsleistung war vor allem von der Intelligenz und der Gewissenhaftigkeit abhängig. Der Beitrag, den die emotionale Intelligenz über die Messungen der akademischen Intelligenz sowie der Persönlichkeit im engeren Sinne hinaus zur Aufklärung der Berufsleistung beitragen konnte, war zwar statistisch bedeutsam, aber absolut gesehen gering. Für Berufe, die „emotionale Arbeit“ erfordern, zeigte sich allerdings eine größere Bedeutung der emotionalen Intelligenz. Bezüglich der Leistungstests (nicht aber der Selbstberichtsverfahren) zeigten sich hinsichtlich der geschlechtsspezifischen Ausprägungen deutliche Leistungsvorteile zugunsten der Frauen.

4 Verfahren der Personalauswahl

Nachdem mit Hilfe der Anforderungsanalyse die relevanten Eignungsmerkmale identifiziert wurden, muss entschieden werden, mit welchen Verfahren Informationen zu diesen Eignungsmerkmalen erhoben werden und wie diese Informationen bewertet und interpretiert werden. Für die Personalauswahl steht eine Vielzahl von Verfahren zur Verfügung, die sich in fünf Gruppen kategorisieren lassen.

Kersting hat ein mehrdimensionales Klassifikationssystem namens CUBE für den gesamten diagnostischen Prozess entwickelt. Mit dem so genannten CUBE 2-1a (vgl. Abb. 3), der in einer Vorform erstmals 2011 publiziert wurde (Kersting, 2011), werden fünf Verfahrenskategorien unterschieden:

- Dokumentenanalyse
- Direkte mündliche Befragungen/Interview
- Verfahren zur Verhaltensbeobachtung und Verhaltensbeurteilung
- Messtheoretisch fundierte Fragebogen
- Messtheoretisch fundierte Tests

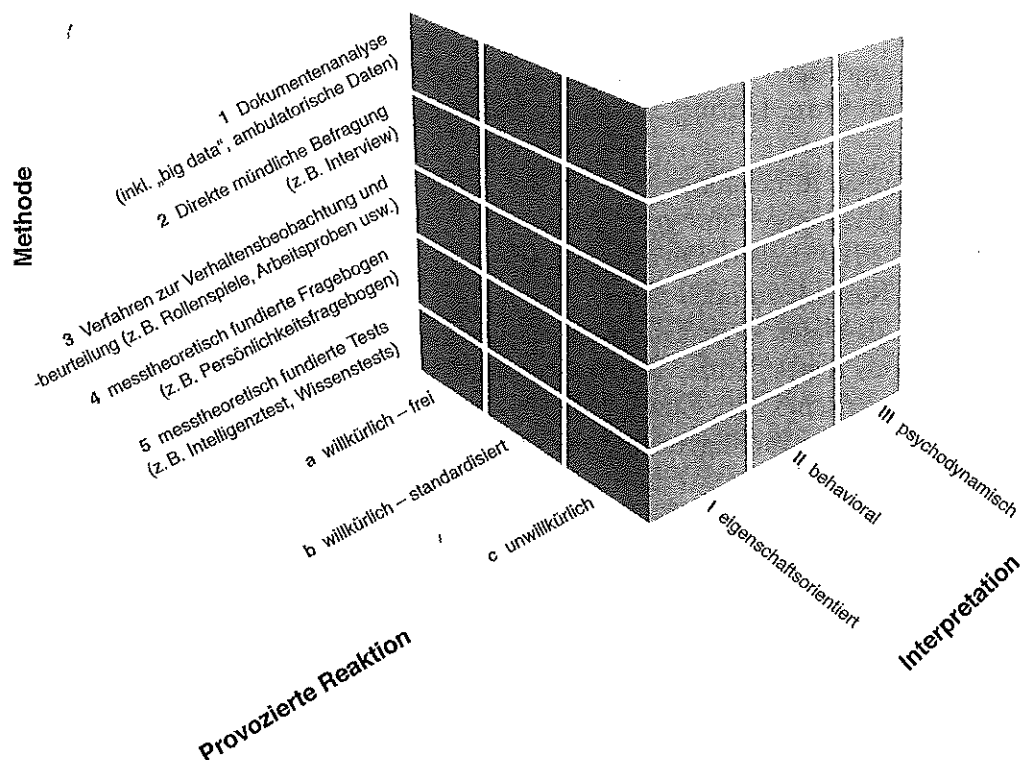


Abbildung 3:

CUBE-System, CUBE 2-1a, Verfahrenskategorien

Eine sechste Kategorie diagnostischer Verfahren stellen psychophysiologische Verfahren dar, die allerdings in der Eignungsdiagnostik keine große Rolle spielen und daher hier unberücksichtigt bleiben. Die oben dargestellte Einteilung in die fünf CUBE Verfahrenskategorien liegt der DIN 33430 (2016) zugrunde und dient uns im Folgenden als Gliederung für die Darstellung der Verfahren.

Aus der Zuordnung der Verfahren zu einer der fünf Kategorien ergeben sich spezifische Qualitätsanforderungen an die Verfahren sowie an die Qualifikation der Personen, die das Verfahren durchführen, auswerten und interpretieren. Die Anforderungen an die Qualifikation der verantwortlichen Personen werden in der DIN 33430 (2016) formuliert und in dem vom Diagnostik- und Testkuratorium (2017) herausgegebenen Buch ausführlich erläutert (vgl. Abschnitt 4.6 in diesem Kapitel).

Welcher Kategorie ein Verfahren zuzuordnen ist, muss im Einzelnen geprüft werden, keinesfalls erfolgt die Zuordnung aufgrund von Oberflächenmerkmalen wie

der Verfahrensbezeichnung. So werden z. B. sowohl messtheoretisch fundierte Tests als auch Verfahren zur Verhaltensbeobachtung und Verhaltensbeurteilung unter der Bezeichnung „Postkorb“ angewendet. In dem einen Fall geht es um vollstandardisierte Computertests (z. B. „PC-Postkorb“ von Funke & Rasche, 1988; zit. n. Funke, 1993; „Mailbox '90“ von Roest, Scherzer, Urban, Gangl & Brandstätter, 1989, oder „PC-Office“ von Fennekels, 1995), in dem anderen Fall bearbeiten Teilnehmer im Rahmen einer AC Übung mehrere Dokumente und sollen anschließend ihre Erkenntnisse und Schlussfolgerungen präsentieren. Ein weiteres Beispiel sind „Arbeitsproben“ (*work sample test*, vgl. Truxillo, Donahue und Kuang, 2004). Unter diesem Sammelbegriff werden unvergleichbare Methoden subsumiert wie high fidelity Simulationen (z. B. Flugsimulatoren), Tests wie die Drahtbiegeprobe (Lienert, 1967) und AC Übungen wie die Gruppendiskussion oder auch die oben genannten Postkörbe. Diese Verfahren stellen jeweils eigene Anforderungen an die Anwender, daher werden sie unterschiedlichen Kategorien zugeordnet (z. B. der Kategorie „messtheoretisch fundierte Tests“ im Falle der Drahtbiegeprobe einerseits und der Kategorie „Verhaltensbeobachtung und Verhaltensbeurteilung“ im Falle der Gruppendiskussion andererseits).

Bevor wir die fünf Verfahrenskategorien näher erläutern, möchten wir kurz die beiden anderen Dimensionen des CUBE 2-1a darstellen. Die zweite Dimension kategorisiert die mit dem Verfahren provozierten Reaktionen der diagnostizierten Personen. Diese können (a) willkürlich-frei sein (reaktionsoffen, wenn Kandidaten z. B. freigestellt ist, wie sie auf Interviewfragen antworten), (b) willkürlich-standardisiert (wenn Kandidaten z. B. in einem standardisierten Interview zwischen vorgegebenen Antwortkategorien wählen müssen) oder (c) unwillkürlich (z. B. bei einem impliziten Assoziationstest).

In der dritten Dimension des CUBE-2-1-a geht es um die Interpretation der Daten. Diese kann (i) eigenschaftsorientiert, (ii) behavioral oder (iii) psychodynamisch erfolgen.

Die Leitidee des CUBE-Systems ist die Mehrdimensionalität. Das CUBE-System zeigt den Möglichkeitsraum diagnostischer Zugänge ungeachtet der Tatsache, dass einige Kombinationen theoretisch nicht möglich sind oder praktisch nicht vorkommen.

Die hier genannten weiteren Dimensionen „provozierte Reaktion“ und „Interpretation“ sind nur Beispiele für eine Menge weiterer Dimensionen des grundsätzlich erweiterungsoffenen CUBE-Systems.

Die übliche eindimensionale Betrachtung von Verfahren führt zu einer Überbetonung der Methoden gegenüber den Inhalten. Dies werden wir im Folgenden

den am Beispiel der offenen Validitätsfragen im Kontext von Interviews herausarbeiten. Es ist ebenso üblich wie fraglich, der aus Interviews abgeleiteten Schlussfolgerungen pauschal eine Validität zu- oder abzusprechen, ohne zu berücksichtigen, was mit dem Interview gemessen wird, wie die gewonnenen Daten interpretiert werden. Die Problematik ergibt sich in verstärkter Form für AC, da hier nicht nur die jeweils erfassten Inhalte variieren, sondern innerhalb eines ACs auch noch unterschiedliche Verfahren im Sinne der CUBE Klassifikation eingesetzt werden.

4.1 Dokumentenanalyse

Bei der Dokumentenanalyse geht es um die zielgerichtete Sichtung und Analyse von Sachinformationen. Im eignungsdiagnostischen Kontext wären hier z. B. Informationen über den schulischen und beruflichen Werdegang zu nennen, die sich u. a. in Bewerbungsunterlagen befinden. Da immer mehr Organisationen den Bewerbungsprozess als online-Verfahren gestalten, liegen Bewerbungsinformationen (z. B. Zeugnisnoten) gehäuft in standardisierter Form vor und sind einer automatischen Auswertung zugänglich. Im weitesten Sinne geht es bei der Dokumentenanalyse um die Analyse von Dokumenten zu „(Verhaltens-)Spuren“, die jemand im Leben hinterlassen hat. Es handelt sich dabei zumeist um „historische“ Daten, so dass bei der Interpretation Annahmen über die Stabilität oder Veränderbarkeit der mit diesen Daten assoziierten Eignungsmerkmale eine Rolle spielen. Einerseits ist das Verhalten und der Erfolg („past merits“) in der Vergangenheit ein guter Prädiktor für das Verhalten und den Erfolg in der Zukunft. Andererseits nimmt die Validität biografischer Daten mit der Zeit ab, d. h. weit zurückliegende biografische Ereignisse haben in der Regel weniger prognostische Aussagekraft.

Unter die Dokumentenanalyse fällt die Auswertung von Bewerbungsunterlagen und Referenzen (vgl. Moser, 2007; Sende & Moser, 2017). Auch Ergebnisse einer Internet-Recherche fallen unter die Dokumentenanalyse, wobei hier besonders die Glaubwürdigkeit der Quelle zu berücksichtigen ist. Grundsätzlich fallen auch „big data“ Analysen in die Kategorie der Dokumentenanalyse, etwa die Analyse von Website-Besuchen, elektronischen Kommunikationsmustern oder Daten von „Wearables“ wie Smart-Glasses (Beermann & Kersting, in Druck). Der Nutzung sind allerdings strenge rechtliche Grenzen gesetzt und die Güte derartiger Analysen ist noch nicht hinreichend untersucht. Insgesamt spielt die Frage der Nachprüfbarkeit eine große Rolle für die Dokumentenanalyse. Außerdem sollte bei der Analyse bedacht werden, ob die Informationen ggf. aus einem privaten Kontext stammen und somit nicht ohne weiteres auf den beruflichen Kontext übertragen werden können.

Die Validität der aufgrund von biografischen Daten getroffenen Interpretationen und Entscheidungen wurde in Einzelstudien nachgewiesen (z. B. Harvey-Cook & Taffler, 2000) sowie in Metaanalysen (Bliesener, 1996) bestimmt. Bliesener (ebd.) schätzte die durchschnittliche prädiktive Validität metaanalytisch auf .22, wobei sich mit biografischen Daten objektive Kriterien (z. B. Verkaufszahlen, Anzahl an Patenten) besser vorhersagen ließen als subjektive Kriterien (z. B. Vorgesetztenurteile). Robertson und Smith (2001) sehen biografische Daten in Bezug auf den Trainingserfolg als zweitbesten Prädiktor (nach Intelligenztests). In einer Studie von Mount, Witt und Barrick (2000) zeigte sich in Bezug auf drei von vier geprüften Kriterien eine inkrementelle Validität von biografischen Daten über die Intelligenztest- oder persönlichkeitsfragebogenbasierten Interpretationen hinaus. Besonders aussagekräftig sind – aller Kritik zum Trotz – Schulnoten. Trapmann, Hell, Weigand und Schuler (2007) fassen die Ergebnisse aus 42 Studien mit insgesamt 45 860 Personen zur Vorhersage des Studienerfolgs durch die Schulnote (Durchschnittsnote) metaanalytisch zusammen und berichten eine Validität von .53 (reliabilitäts- und selektionskorrigierter Wert). In einer älteren Metaanalyse wurde die Vorhersagekraft der Schulnoten für den Ausbildungserfolg auf .41 geschätzt (Baron-Boldt, Funke & Schuler, 1989). Inhaltlich ist davon auszugehen, dass die Schulnoten einen Indikator für die Intelligenz darstellen. Roth und Kollegen (2015) schätzen den Zusammenhang zwischen Schulnoten und Intelligenz anhand einer Metastudie mit Daten von insgesamt rund 100 000 Personen auf .54. Die höchsten Zusammenhänge zeigten sich dabei zwischen Intelligenzleistungen und Schulleistungen in Mathematik und Naturwissenschaften.

4.2 Direkte mündliche Befragungen und Interviews

Mit direkten mündlichen Befragungen sind nach der DIN 33430 (2016) sowohl Interviews mit Bewerbern als auch Gespräche mit Referenzgebern oder anderen relevanten Auskunftspersonen gemeint. Das Gespräch kann persönlich geführt werden, aber auch technisch vermittelte Gespräche (Telefon, Video usw.) fallen in diese Kategorie. Das Interview ist – nach der (Vor-)Auswahl aufgrund von Bewerbungsunterlagen – das Verfahren, das in der externen Personalauswahl am häufigsten zur Anwendung kommt. Zur Verbreitung eignungsdiagnostischer Verfahren zur externen Personalauswahl im deutschsprachigen Raum vgl. z. B. Schuler, Hell, Trapmann, Schaar & Boramir, 2007. Die entsprechenden Befunde zur internen Personalauswahl – und somit indirekt zur Leistungsbeurteilung – wurden von Hell, Schuler, Boramir und Schaar (2006) publiziert.)

Ungeachtet der methodischen Schwächen (s. u.) und des großen Aufwands erfreut sich das Interview sowohl bei den Betroffenen (vgl. z. B. Anderson, Salgado & Hülsheger, 2010; Hausknecht, Day & Thomas, 2004) als auch bei den

Interviewern (vgl. Schuler et al., 2007) einer ungebrochenen Beliebtheit. Schuler (1992, S. 282) sieht hier in Bezug auf konventionell geführte Interviews eine „bemerkenswerte Diskrepanz zwischen subjektiver Wertschätzung und empirischer Bewertung“. Diese Diskrepanz erklärt sich womöglich u. a. aus der Multifunktionalität des Interviews (vgl. z. B. Dipboye, 1989, S. 60). Während andere Verfahren wie messtheoretisch fundierte Fragebogen oder Tests einseitig der Informationsermittlung der Organisation dienen (gleichwohl über ein Feedback auch die Möglichkeit zur Information der diagnostizierten Personen besteht), kommt dem Interview – häufig entgegen seiner expliziten Zielvorgabe – auch informationsvermittelnde Funktion zu. Im Multimodalen Interview (Schuler, 1992) ist explizit eine Phase zur Information über die Tätigkeit vorgesehen. Diese Gestaltung trägt der Tatsache Rechnung, dass im Interview meistens auch Informationen über die einstellende Organisation und die Tätigkeit gegeben werden und Bewerber die Möglichkeit haben, Fragen zu stellen. Die explizite Berücksichtigung dieses Interviewinhalts in einer definierten, nicht bewertungsrelevanten Phase trägt dazu bei, dass nur anforderungsrelevante Informationen verwertet werden und nicht etwa die vermeintliche Begeisterung des Bewerbers angesichts des Berichts des Interviewers über seine Heldentaten bei der Unternehmensgründung.

Ein weiterer Grund für die Wertschätzung, die Interviewte dem Interview zukommen lassen, dürfte in der unmittelbaren (verbalen und nonverbalen) Reaktion des Interviewers liegen. Die Teilnehmer erhalten in einem Interview das Feedback von den Entscheidern, welches sie „als Objekt der Beobachtung oder Bewertung“ bei der Bearbeitung von Fragebogen und Tests oder als Teilnehmer einer Gruppendiskussion vermissen. Schließlich erlaubt das Interview den Interviewten eine optimale Selbstpräsentation, die sie in stärker standardisierten Verfahren so nicht erleben können (e. g., Stevens & Kristof, 1995). In dieser Hinsicht nimmt das Interview im Rahmen der Personalauswahl- und -beurteilungsverfahren aufgrund der unmittelbaren Interaktion zwischen Beurteilendem und Beurteilten eine Sonderstellung ein. Es unterscheidet sich von anderen Verfahren dadurch, dass es auf einen echten Dialog hin ausgelegt ist. Dieser Dialogcharakter wirkt sich ungünstig auf die Objektivität und Reliabilität, aber günstig auf die Akzeptanz aus. Eine stärkere Standardisierung der Interviewsituation, die letztendlich in einer Angleichung an Fragebogen gipfelt, indem die interviewten Personen zwischen vorgegebenen Antworten wählen können, geht zu Lasten der Akzeptanz. Die dialogische Situation impliziert eine aktive Teilnahme der Interviewer, unter Gesichtspunkten der Verhaltensbeobachtung handelt es sich um eine teilnehmende (im Gegensatz zu einer nicht teilnehmenden) Beobachtung. Die aktive Teilnahme in Form von Fragen bindet die Aufmerksamkeit der Interviewer und erschwert die Beobachtungs- und Beurteilungstätigkeit (Doppelbelastung). Häufig wird als Gegenmaßnahme ein zwei- oder mehrköpfiges Team eingesetzt (Panel oder board interviews), das mit „verteilt-

ten Rollen“ arbeitet (z. B. unterschiedliche Aktivitätsgrade pro „Phase“). Ungeachtet der Anzahl der Interviewer wird das Verhalten der beobachteten und beurteilten Personen durch verbale und nonverbale Reaktionen des Interviewers oder des Interviewerteams verändert. Umgekehrt wird das Urteil der Interviewer durch nonverbale Reaktionen des Interviewten beeinflusst (Levine & Feldmann, 2002; vgl. auch Dipboye, 1992).

Die Frage nach der Kriteriumsvalidität von Interviews scheint auf den ersten Blick sehr gut und mit befriedigenden Ergebnissen geklärt. Bereits Schmidt und Hunter (1998) berichten Kriteriumsvaliditäten von $r = .38$ und $r = .51$ für Eignungsaussagen, die aus unstrukturierten oder strukturierten Interviews abgeleitet werden. Moscoso (2000) kann in ihrer Übersichtsarbeit auf acht Metaanalysen zurückblicken, die ausnahmslos zu dem Ergebnis kommen, dass sich aus Interviews kriteriumsvalide Aussagen ableiten lassen. Dabei gelten strukturierte Interviews als aussagekräftiger als unstrukturierte Interviews. In der Metaanalyse von McDaniel, Whetzel, Schmidt und Maurer (1994), in die insgesamt 245 Validitätskoeffizienten einbezogen wurden, die an 86 311 Personen gewonnen wurden, ermittelten die Autoren für das strukturierte Interview einen Wert von $.44$ und für das unstrukturierte Interview einen Wert von $.33$.

In einer Metaanalyse schätzten Huffcutt, Culbertson und Weyhrauch (2014) – wenn auch unter Nutzung extremer Korrekturen für Varianzeinschränkung – die Kriteriumsvalidität von Eignungsaussagen, die sich aus Interviews ableiten, sogar auf $.71$ für das zweithöchste und $.70$ für das höchste Strukturierungslevel.

Innerhalb der Gruppe der strukturierten oder unstrukturierten Interviews variieren die Validitäten in Abhängigkeit von dem genutzten „Interviewsystem“. Zu den strukturierten Interviews zählen das biografische und das situative Interview. In einer Studie von Klehe und Latham (2005) erzielten Vorhersagen zu „Teamplayer-Leistungen“ auf der Grundlage des situativen Interviews nicht nur nominell höhere Validitäten als die Vorhersagen auf der Basis eines biografischen Interviews ($.41$ zu $.34$), sondern das situative Interviewsystem erzielte auch inkrementelle Validität über den Beitrag des biografischen Interviewsystems hinaus. Nach Huffcutt, Conway, Roth und Klehe (2004) muss auch die validitätsmoderierende Rolle der Komplexität der jeweils in Frage stehenden Berufe berücksichtigt werden. Während sich in der Studie von Huffcutt und Kollegen (ebd.) situative Interviews für wenig bis mittel komplexe Berufe gut eigneten ($.44$ und $.51$), nahm die Validität mit zunehmender Jobkomplexität ab ($.30$ für komplexe Berufe).

Trotz der großen Menge an Daten und (Meta-)Analysen ist es auf den zweiten Blick nicht leicht, ein klares Bild von der Validität interviewbezogener Interpretationen und Entscheidungen zu gewinnen. Dies hat methodische und inhaltliche Gründe. Zunächst ist die in der Literatur gängige Unterscheidung zwischen

strukturierten und unstrukturierten Interviews in mehrfacher Hinsicht zu hinterfragen. Für ein vollständig frei geführtes Interview existieren keine Daten, die in Analysen eingehen könnten. In der Praxis führen Interviewer häufig mit mehreren Bewerbern ein freies Interview und stellen eine der Personen ein. Aufzeichnungen oder gar formale, skalierte Leistungsbewertungen über das Abschneiden der Personen im Interview liegen nicht vor. Diese Interviews, die die Realität eines vermutlich großen Teils der Praxis ausmachen, kommen in wissenschaftlichen Analysen mangels Datengrundlage nicht vor. Als „unstrukturiert“ werden in der Fachliteratur hingegen Interviews klassifiziert, die lediglich weniger Strukturierungsaspekte aufweisen als die als „strukturiert“ klassifizierten Interviews. Aus dieser Tatsache lässt sich schlussfolgern, dass die Validität von Entscheidungen auf der Basis unstrukturierter Interviews in der Fachliteratur überschätzt wird.

Auf der anderen Seite ist davon auszugehen, dass die Metaanalysen in Bezug auf die strukturierten Interviews einen „publication bias“ aufweisen (Duval, 2005), was zu einer Überschätzung der Validität strukturierter Interviews führen dürfte. Der Effekt ist so erheblich, dass der Validitätsvorteil für strukturierte Interviews in der Metaanalyse von McDaniel und Kollegen (1994) nach Korrektur des Bias beträchtlich schrumpfte. Kepes, Banks, McDaniel und Whetzel (2012) bestätigen den „publication bias“, gehen aber davon aus, dass die Schlussfolgerungen der Metaanalyse auch nach Korrektur des Bias zutreffend sind.

Neben diesen Repräsentationsproblemen bezüglich der Stichprobe im Allgemeinen (keine Berücksichtigung wirklich unstrukturierter Interviews) und im Besonderen (publication bias), gibt es weitere methodische Bedenken. Schmidt und Zimmerman (2004) geben zu bedenken, dass die höhere Validität strukturierter Interviews möglicherweise lediglich auf deren höhere Reliabilität zurückzuführen sei. Auch in diesem Fall müssten die Befunde methodisch und nicht inhaltlich interpretiert werden. Allerdings werden in Metaanalysen in der Regel reliabilitätskorrigierte Werte berichtet.

Das Hauptproblem besteht aber darin, dass nicht einheitlich definiert ist, was überhaupt ein strukturiertes Interview ist und wie man den Grad der Strukturiertheit operationalisiert. Vorschläge wurden unter anderem von Chapman und Zweig (2005); Dipboye, Wooten und Halverson (2004) sowie Huffcutt und Kollegen (2014) entwickelt, keinesfalls folgt die vorhandene Literatur aber einheitlich diesen Einteilungsvorschlägen. Huffcutt und Kollegen (2014) schlagen vor, zwischen dem Grad der Standardisierung der Fragen und dem Grad der Standardisierung der Bewertung der Antworten zu unterscheiden, Dipboye und Kollegen (2004) sehen den Anforderungsbezug der Fragen als weiteren Einteilungsgesichtspunkt.

Inhaltlich kranken die Metaanalysen zur Kriteriumsvalidität von Interviews daran, dass als Ordnungsgesichtspunkt lediglich die Methode im Allgemeinen

(Interview) oder Besonderen (Strukturierung, Interviewsystem) herangezogen wird, der Inhalt des Interviews aber unberücksichtigt bleibt. Die vorrangige Frage bleibt die nach der Konstruktvalidität des Interviews. Häufig weiß man nicht, auf welche Konstrukte die Interviews abzielten, die in den Validitätsstudien berücksichtigt wurden. Im Falle von strukturierten Interviews sollten dies – zumindest nach etablierten Definitionen der Strukturierung – anforderungsrelevante Dimensionen gewesen sein. Selbst wenn – was nicht immer der Fall ist – eine Information über den Anforderungsbezug der Fragen vorliegt, bleibt unklar, ob die als anforderungsrelevant bezeichneten Konstrukte tatsächlich anforderungsrelevant waren (weil die Validität von Anforderungsanalysen, wie bereits oben ausgeführt, nicht geprüft wird, vgl. Kersting & Birk, 2011), noch ob es gelungen ist, mit dem Interview die intendierten Konstrukte zu erfassen.

Huffcutt, Conway, Roth und Stone (2001) reanalysierten 47 Studien um herauszufinden, welche Konstrukte mit einem Interview erfasst wurden. Demzufolge werden Interviews insbesondere zur Erfassung der Persönlichkeit (weitestgehend im Sinne des weiter oben thematisierten Fünf-Faktoren-Modells, ergänzt um biografische Informationen wie einschlägigen Erfahrungen) und zur Erfassung sozialer Kompetenzen eingesetzt (35 % und 28 %). Weitere Konstrukte sind geistige Fähigkeiten (16 %) und berufliches Wissen sowie berufliche Fertigkeiten (10 %). Bezüglich der Dimensionen innerhalb einer Kategorie variiert die Häufigkeit, mit der ein Konstrukt durch ein Interview erfasst wird, in Abhängigkeit vom Strukturierungsgrad. Strukturierte Interviews erfassen der Analyse von Huffcutt und Kollegen (ebd.) zufolge eher angewandte geistige Fähigkeiten wie Problemlösen, Entscheiden und Planen, während unstrukturierte Interviews eher allgemeine Intelligenz erfassen. Innerhalb der Kategorie Persönlichkeit werden mit unstrukturierten Interviews eher Merkmale wie emotionale Stabilität und Verträglichkeit erfasst, während strukturierte Interviews eher auf die Offenheit für Erfahrung abzielen. Insgesamt werden mit strukturierten Interviews – vermutlich in Folge der anforderungsanalytisch fundierten Entwicklung dieser Interviews – eher (job-)spezifischere Konstrukte erfasst als mit unstrukturierten Interviews. Anforderungsanalytisch orientierte Interviews erlauben aber im Allgemeinen eher valide Entscheidungen (Wiesner & Cronshaw, 1988). Die Befunde konvergieren weitgehend mit den metaanalytischen Befunden von Salgado und Moscoso (2002), die konventionelle Interviews mit verhaltensbasierten Interviews verglichen. Während mit konventionellen Interviews eher allgemeine Intelligenz (.41) und allgemeine Persönlichkeitsmerkmale wie emotionale Stabilität (.38) erfasst werden, zielen verhaltensbasierte Interviews auf Berufserfahrung (.71), berufsbezogenes Wissen (.53) und soziale Kompetenzen (.65). Offensichtlich erfassen die unterschiedlichen Interviewformen unterschiedliche Konstrukte. Gleichzeitig weisen die Konstrukte Huffcutt und Kollegen (2001) zufolge unterschiedliche Validitäten auf von .58 (Kreativität) bis .24 (allgemeine Intelligenz sowie Interessen). Diese Beobachtung stellt den Be-

fund des Validitätsvorteils strukturierter Interviews abermals in Frage. Zumindest ist unklar, ob das Validität-generierende Moment die Struktur des Interviews ist oder die Tatsache, dass die Zielkonstrukte anforderungsanalytisch festgelegt wurden – was nur ein Definitionsbestandteil strukturierter Interviews ist. In vielen Metaanalysen zur Kriteriumsvalidität von Interviews sind Methoden und Konstrukte konfundiert.

4.3 Verfahren zur Verhaltensbeobachtung und Verhaltensbeurteilung

Unter die Kategorie „Verfahren zur Verhaltensbeobachtung und Verhaltensbeurteilung“ fallen in der CUBE Systematik und somit in der DIN 33430 (2016) Aufgaben wie Rollenspiele, Gruppendiskussionen, Präsentationen und Fallstudien. Solche Verfahren werden eingesetzt, um gezielt Verhalten hervorzurufen, welches dann beobachtet und beurteilt wird. Sofern die Beurteilung von Konstrukten intendiert ist, soll die Situation das Merkmal im Sinne Murrays (1938) zum Ausdruck bringen. Die Situationen können diesbezüglich unterschiedlich stark oder schwach sein (Mischel, 1977). Eine differenzierte Beschreibung unterschiedlicher Situationen und deren Interaktion mit Persönlichkeitsmerkmalen im Arbeitskontext haben Tett und Guterman (2000) sowie Tett und Burnett (2003) im Rahmen ihrer Trait Activation Theorie vorgestellt.

Verfahren zur Verhaltensbeobachtung und Verhaltensbeurteilung werden – normalerweise unter dem Begriff „Übungen“ oder „Stationen“ – häufig im Rahmen eines Assessment-Centers (AC) oder Development-Centers (DC) eingesetzt. Die Übungen des ACs sollen eine Stichprobe der realen beruflichen Aufgaben darstellen. Idealerweise bilden die einzelnen AC-Übungen den zu prognostizierenden Leistungsbereich ab.

Mit „Verfahren zur Verhaltensbeobachtung und Verhaltensbeurteilung“ ist aber nicht das AC gemeint. Ein AC ist ein Methodenmix, d. h. in einem AC kommen mehrere Verfahren zum Einsatz, die mehreren, unterschiedlichen Kategorien des CUBE-Systems zuzuordnen sind. So ist beispielsweise ein Interview (77 %) häufig Bestandteil eines ACs, auch die Bearbeitung von messtheoretisch fundierten Fragebogen (43 %) und Tests (40 %) steht häufig auf dem AC Ablaufplan, Dokumentenanalysen im Sinne biografischer Fragebogen sind seltener (9 %) anzutreffen, aber auch nicht ausgeschlossen. In Klammern ist jeweils angegeben, wie häufig die genannten Verfahren der AC Studie des Arbeitskreises AC 2012 zufolge im Rahmen eines ACs eingesetzt werden (Höft, Obermann & Becker, 2014). An der Studie haben Vertreter von 125 Unternehmen (darunter alle DAX Unternehmen) teilgenommen.

Verfahren der Verhaltensbeobachtung und -beurteilung sind also ein (sogar konstituierender) Bestandteil von ACn, keinesfalls aber gleich zu setzen mit ACn. Bei Verfahren der Verhaltensbeobachtung und -beurteilung geht es um eine geplante Beobachtung und Protokollierung direkten Verhaltens mit dem Ziel, Eignungsaussagen abzuleiten.

Ungeachtet der Multimethodalität befasst sich ein Großteil der Forschung mit der Frage, wie valide Interpretationen sind, die aus ACn abgeleitet werden. Solche Aussagen sind schwer zu bewerten, da nicht nur – wie beim Interview – die jeweils erfassten Konstrukte variieren, sondern auch die Methoden (z. B. Verhaltensbeobachtung vs. Testauswertung).

Bezüglich der mit einem AC erfassten Konstrukte haben wir weiter oben bereits sowohl die Arbeit von Arthur und Kollegen (2003) als auch die Arbeit von Gibbons und Kollegen (2006) erwähnt, die auf der Basis von empirischen Studien ermittelt haben, welche Dimensionen häufig mit einem AC erfasst werden. Merriac, Hoffman und Woehr (2014) gelangen in konzeptioneller Fortschreibung dieser Arbeiten und auf der Basis einer Kombination von theoretischem und empirischem Vorgehen zu einer Dreier-Struktur. Demzufolge werden im AC auf der generellen Ebene „administrative skills“ (Planen, Problemlösen usw.), „relational skills“ (Kommunikation, Führung usw.) sowie „drive“ (Initiative, Motivation) erfasst. Wie zu erwarten sind die administrativen Skills enger mit der kognitiven Kompetenz verbunden als die relationalen Skills, wenngleich sich für alle AC-Dimensionen substantielle Verbindungen zur Intelligenz (sowie zur Extraversion) aufzeigen lassen. Allerdings gilt grundsätzlich, dass es den aus Assessment Centern abgeleiteten Eignungsaussagen an Konstruktvalidität mangelt (vgl. z. B. Bowler & Woehr, 2006). Theoretisch sollen die einzelnen Konstrukte (z. B. Kommunikation) in verschiedenen Übungen, z. B. in der Gruppendiskussion und im Rollenspiel erfasst werden. Die Analysen zeigen hingegen immer wieder, dass die Assessoren eher Urteile über das Verhalten pro Übung abgeben und die Konstrukte innerhalb einer Übung nicht hinreichend differenziert beurteilen.

Zur Erfassung von Konstrukten oder eines Gesamteindrucks (overall assessment center rating, OAR) werden im AC Informationen aus verschiedenen Übungen oder Stationen berücksichtigt. Die berücksichtigten AC-Übungen gehören, wie oben dargestellt, keineswegs alle zur Kategorie Verhaltensbeobachtung und Verhaltensbeurteilung im Sinne der CUBE Systematik. Es handelt sich beispielsweise um Interviews (vgl. Abschnitt 4.2 in diesem Kapitel) oder um messtheoretisch fundierte Fragebogen oder Tests (vgl. Abschnitt 4.4 und 4.5 in diesem Kapitel). Eine Studie, die sich auf Verfahren der Verhaltensbeobachtung und Verhaltensbeurteilung im AC konzentriert, wurde von Hoffman, Kennedy, Lopilato, Monahan und Lance (2015) vorgestellt. Die Autoren untersuchen u. a. metaanalytisch die Konstrukt- und Kriteriumsvalidität der aus fünf gängigen

AC-Übungstypen abgeleiteten Interpretationen. Bei diesen Übungen handelt es sich um (1) den Postkorb, (2) die führerlose Gruppendiskussion (im Folgenden als „Gruppendiskussion“ bezeichnet), (3) das Rollenspiel, (4) die Fallstudie und (5) die Präsentationsübung.

Während Schmidt und Hunter (1998) eine Kriteriumsvalidität von $r = .37$ für die aus dem Gesamt-ACn (OAR) abgeleiteten Eignungsaussagen berichten, liegen die Kriteriumsvaliditäten der aus den fünf Übungstypen abgeleiteten Eignungsaussagen der Studie von Hoffmann und Kollegen zufolge im Bereich von .16 bis .19. Damit sind die für die Einzelübungen im AC berichteten Validitäten geringer als die Validitäten, die in anderen Studien für AC Dimensionen oder Gesamtratings berichtet werden. Dieser Befund kann so interpretiert werden, dass die Validität der aus Verhaltensbeobachtungen und -beurteilungen abgeleiteten Eignungsaussagen vergleichsweise geringer ist als die Validität der aus anderen Assessment-Center Bestandteilen abgeleiteten Eignungsaussagen. Anders ausgedrückt: Die Validität AC-gestützten Eignungsaussagen wird nicht durch, sondern trotz der Übungen aus dem Bereich Verhaltensbeobachtungen und -beurteilungen erzielt. Hierzu passt der Befund der Metaanalyse von Becker, Höft, Holzenkamp und Spinath (2011) zur Kriteriumsvalidität der aus ACn abgeleiteten Eignungsaussagen. Hinsichtlich der inhaltlichen Gestaltung erwiesen sich AC insbesondere dann als aussagekräftig, wenn im Rahmen des ACs ein Intelligenztest eingesetzt wurde (.56). Der Verzicht auf den Einsatz von Intelligenztests führte zu einem Absinken der Validität auf $r = .25$.

4.4 Messtheoretisch fundierte Fragebogen

Unter die Kategorie messtheoretisch fundierter Fragebogen fallen alle standardisierten Selbstberichtsverfahren. Im Kontext der Personalauswahl werden besonders häufig Persönlichkeitsfragebogen eingesetzt, beispielsweise zur Erfassung der Dimensionen und Facetten des Fünf-Faktoren-Modells (FFM) (vgl. Abschnitt 3.2.4 in diesem Kapitel). Weitere Verfahren sind Motivations- und Interessensfragebogen (vgl. Abschnitt 3.2.2 und 3.2.3 in diesem Kapitel). Über die Validität der aus diesen Verfahren abgeleiteten Interpretationen haben wir bereits in den genannten Abschnitten berichtet. Die Befunde sind, wie oben bereits erwähnt, eindeutiger als die Befunde zur Validität der Interpretationen, die aus Verhaltensbeobachtungen/-beurteilungen und/oder Interviews abgeleitet werden. Dies liegt daran, dass in die entsprechenden Studien überwiegend nur Fragebogen-Daten eingegangen sind und eine nach Konstrukten (z. B. Dimensionen des FFM) differenzierte Auswertung vorgenommen wurde, so dass eine identifizierbare Einheit aus Konstrukt einerseits und einer Methode andererseits vorliegt. Im Kontext der Personalauswahl ist der Anforderungsbezug in mehrfacher Hinsicht von besonderer Bedeutung. Persönlichkeitsfragebogen erlauben

validere Interpretationen, wenn damit Dimensionen erfasst werden, die sich aufgrund von Anforderungsanalysen als bedeutsam erwiesen haben (Tett, Jackson & Rothstein, 1991), wobei hier die Validität der Anforderungsanalyse wiederum vorausgesetzt werden muss. Auch ein Anforderungsbezug der im Persönlichkeitsfragebogen genutzten Items wirkt sich günstig auf die Validität der Interpretationen aus (Bowling & Burns, 2010; Reddock, Biderman & Nguyen, 2011).

Die kontextuelle Einbettung der Items eines Persönlichkeitsfragebogens dürfte sich auch positiv auf die Augenscheinvalidität und somit auf die soziale Akzeptanz (Schuler, 1990; Hausknecht et al., 2004) auswirken. Diesem Nebengütekriterium kommt für Persönlichkeitsfragebogen eine hohe Bedeutung zu. In einer Metaanalyse (Anderson et al., 2010) erzielten Persönlichkeitsfragebogen in Bezug auf die Gesamtakzeptanz zwar positive Werte, landeten nominell aber auf Platz 7 von 10 möglichen Plätzen, weit hinter dem Interview und nominell auch hinter Intelligenztests. Dies könnte erklären, warum Persönlichkeitsfragebogen vergleichsweise selten eingesetzt werden (Schuler et al., 2007).

4.5 Messtheoretisch fundierte Tests

Das prominenteste Beispiel für ein Verfahren der Kategorie der messtheoretisch fundierten Tests ist der Intelligenztest. Wie bereits oben dargestellt, erlauben Interpretationen auf der Basis von Intelligenztests die mit Abstand validesten Interpretationen im Kontext der Personalauswahl. Die Studien lassen sich aufgrund der Einheit von Methode und Konstrukt eindeutig interpretieren, obgleich auf Konstruktebene häufig nicht nach einzelnen Intelligenzdimensionen differenziert wird. Da Intelligenztests allerdings positiv miteinander korrelieren (positive manifold), wird die Interpretierbarkeit der Befunde durch die mangelhafte Differenzierung nur vergleichsweise gering eingeschränkt.

Unter die Kategorie der messtheoretisch fundierten Tests fallen alle Verfahren mit einem vorab festgelegten „richtig“/„falsch“ Schema. Neben den Intelligenztests (einschließlich der Wissens- und Konzentrationstests) fallen die im Text bereits erwähnten Tests zu praktisch-mechanisch-räumlich-physikalischen Fähigkeiten und die Leistungstests zur sozialen oder emotionalen Intelligenz sowie die SJTs und Problemlöseszenarien in die Kategorie der messtheoretisch fundierten Tests.

Ein gemeinsames Element der Verfahren dieser Kategorie ist, dass die Leistungserbringung Anstrengung und somit Motivation voraussetzt. Eine geringe Leistung kann somit sowohl auf eine gering ausgeprägte Fähigkeit als auch auf mangelnde Anstrengungsbereitschaft zurückgeführt werden. Die Motivation spielt nicht nur während, sondern ggf. auch schon vor der Testung eine Rolle.

Tests, die im Personalauswahlkontext eingesetzt werden, können vorab geübt werden. Für die messtheoretisch fundierten Tests sind die Effekte der Vorbereitung und Übung besonders gut untersucht, sie haben sich als eher gering erwiesen. Als Resultat einer Metaanalyse über 50 Studien berichten Hausknecht, Halpert, Di Paolo und Moriarty Gerrard (2007) eine Effektgröße von .26. Deutlichere Effekte (.46) zeigten sich für den Fall, dass identische Tests wiederholt eingesetzt werden. Problematisch ist der Effekt vor allem dann, wenn nur einige Teilnehmer vorab üben oder sich vorbereiten. Daher empfiehlt es sich, allen Testteilnehmern vorab Übungsmaterial zur Verfügung zu stellen.

Ein weiteres relevantes Thema im Kontext der messtheoretischen Tests ist die mit Leistung im Allgemeinen oder mit Leistungstests im Besonderen verbundene Angst. Metaanalytische Studien zeigen einen (geringen) Zusammenhang zwischen Neurotizismus und allgemeiner Intelligenz (.06 für emotionale Stabilität, Poropat, 2009). Für die Testangst liegen Erkenntnisse aus älteren Studien vor (Hembree, 1988), demzufolge geht Besorgtheit (-.31) bzw. Aufgeregtheit (-.15) mit schlechteren kognitiven Leistungen einher (Hembree, 1988). Metaanalytisch konnte gezeigt werden, dass Testangst zu -.24 mit den „grade point average“ (GPA) Leistungen einhergeht (Richardson, Abraham & Bond, 2012).

Die größten Bedenken gegen den Einsatz von messtheoretisch fundierten Tests, insbesondere gegen den Einsatz von Intelligenztests, liegen im Bereich der sozialen Akzeptanz. Dabei landeten Intelligenztests in der Metaanalyse von Anderson und Kollegen (2010) hinsichtlich ihrer Akzeptanz im mittleren Bereich, nominell vor den Persönlichkeitsfragebogen. Intelligenztests belegten sogar den ersten Platz, was die Wahrung der Privatsphäre angeht. Auch Kersting (2008) zog nach seiner empirischen Analyse der Akzeptanz von fünf gängigen Intelligenztests eine positive Bilanz. Über alle Tests hinweg wurde die Messqualität und vor allem die Kontrollierbarkeit von Intelligenz- und Leistungstests besonders positiv beurteilt. Relativ kritischer wurde hingegen die Augenscheinvalidität der Verfahren beurteilt, wobei hier insbesondere Verfahren mit nur einem Aufgabentyp (APM, d2) schlecht abschnitten. Insgesamt gelten Intelligenztests eher als abstrakt und realitätsfern.

4.6 Verfahrenübergreifende Gesichtspunkte sowie Verfahrenskombinationen

Neben den spezifischen Vor- und Nachteilen der einzelnen Verfahrenskategorien gibt es Themen, die für mehrere Verfahrenskategorien bedeutsam sind. Insbesondere die direkten mündlichen Befragungen (Interviews) sowie die Ver-

fahren der Verhaltensbeobachtung und -beurteilung stellen umfassende Anforderungen an die durchführenden Personen, die quasi das Messinstrument darstellen. Nach DIN 33430 (2016) benötigt man für beide Verfahrenskategorien Kenntnisse und Erfahrungen in der Verhaltensbeobachtung und -beurteilung. Die entsprechenden Qualifikationsanforderungen werden in der DIN 33430 (2016) benannt und von Höft und Kersting (2017) erläutert. Für die Durchführung und Auswertung von direkten mündlichen Befragungen oder Interviews bedarf es nach DIN 33430 (2016) zusätzlich weiterer Kenntnisse. Beispielsweise sind dazu nach DIN 33430 (2016) Kenntnisse über die Handhabung von Interviewleitfäden notwendig, über Frage- und Formulierungstechniken sowie über die rechtliche Zulässigkeit von Fragen (vgl. die Erläuterungen der diesbezüglichen DIN-Anforderungen bei Strobel, Franke-Barthold & Kersting, 2017).

Darüber hinaus sind beide Verfahrenskategorien im besonderen Maße davon geprägt, dass die diagnostizierten Personen mit Techniken der Selbstdarstellung (Impression Management) das Verfahrensergebnis beeinflussen können. Auch im Bereich der Leistungsbeurteilung (vgl. Abschnitt 6 in diesem Kapitel) spielt das Thema bei bestimmten Beurteilungsmethoden eine Rolle, weil Mitarbeiter versuchen können, einen „guten Eindruck“ von sich zu vermitteln. McFarland, Yun, Harold, Viera und Moore (2005) unterscheiden diesbezüglich selbstförderliche Techniken, defensive Techniken (zum Schutz des aufgebauten Images) und nonverbale Taktiken.

Auch bei messtheoretisch fundierten Fragebogen ist es möglich, einen „guten Eindruck“ von sich zu vermitteln.

Die Fähigkeit zur Selbstdarstellung ist in vielen Berufen allerdings erfolgsrelevant, so dass die Beeinflussbarkeit eines Verfahrens durch Selbstdarstellung keinen grundsätzlichen Nachteil in der Kriteriumsvalidität nach sich zieht. Barrick, Shaffer und DeGrassi (2009) konnten beispielsweise zeigen, dass die Selbstdarstellung im Interview mit dem Berufserfolg korreliert. Eine positive Selbstdarstellung setzt voraus, dass man die jeweiligen Anforderungen erkennt und das eigene Verhalten darauf abstimmt – auch diese Fähigkeiten sind berufserfolgsrelevant (Melchers, Ingold, Wilhelmy & Kleinmann, 2015). Von daher gefährdet die Selbstdarstellung nicht grundsätzlich die Kriteriumsvalidität der aus Interviews, Verhaltensbeobachtungen und Persönlichkeitsfragebogen abgeleiteten Eignungsaussagen – sie stellt aber die Konstruktvalidität der Interpretationen in Frage.

In der Regel wird man eine Kombination aus verschiedenen Verfahrenskategorien einsetzen. Die Kombination ist u. a. unter den Gesichtspunkten der Validität, Effizienz und Akzeptanz zu gestalten.

Viele Verfahren können elektronisch vermittelt eingesetzt werden, teilweise ist es möglich, die Verfahren ohne Aufsicht durchzuführen. Was dabei zu beachten ist, steht u. a. in den entsprechenden Richtlinien der International Test Commission (ITC, 2006).

5 *Arbeitsleistung*

Mit den Ergebnissen der Verfahren der Personalauswahl will man letztlich vorhersagen, ob Kandidaten für eine Stelle geeignet oder nicht geeignet sind. Aus Organisationsperspektive läuft dies häufig auf die Frage hinaus, wie gut die Leistung der entsprechenden Kandidaten auf der betreffenden Stelle ausfallen wird. Die Leistung von Mitarbeitern ist (neben Zufriedenheit, Gesundheit und Stress) eine der tragenden Ergebnisvariablen der Arbeits- und Organisationspsychologie, sie ist aber leider auch eine Variable, die nicht leicht zu fassen ist.

Binning und Barrett (1989) bieten eine recht weitreichende Definition von Leistung als Verhaltens- und Ergebnisdomäne, welche als Maßstab für die Auswahl und/oder Entwicklung bestimmter Leistungsvorhersage- und -bewertungssysteme dient. Einfach gesagt: Leistung ist hier das Verhalten oder Ergebnis, was wir im jeweiligen Kontext als erstrebenswertes Kriterium „Leistung“ definieren und daraufhin zu messen und vorherzusagen versuchen. Ein Problem dieser weit gefassten Definition ist jedoch, dass gerade Ergebnisse (Wie viele Stücke wurden produziert? Wie viel Gewinn im Verkauf erzielt? etc.) häufig zwar besonders relevant für die Organisation sind, jedoch nur bedingt der Kontrolle der leistenden Person unterliegen. Produktionsraten hängen beispielsweise auch von der Zulieferung entsprechender Rohstoffe und der technischen Ausstattung ab, der Verkauf hängt von der gesamtwirtschaftlichen Lage, der Konkurrenz vor Ort, der Attraktivität des Produktes und Effektivität von Marketingkampagnen etc. ab.

Aus diesem Grund folgen die meisten Psychologen der Argumentation Campbells (1990), der Leistung enger definiert als beobachtbares Verhalten eines Mitarbeiters, welches für die Zielerreichung der Organisation relevant ist. Die Ergebnisdomäne dagegen ist bei Campbell nicht Teil von Leistung, sondern nur ein Ergebnis davon.

5.1 Rahmenmodell

Wenn wir nun also Leistung als das Verhalten von Mitarbeitern definieren, das für die Zielerreichung der Organisation direkt oder indirekt relevant ist, so sollten wir das Konstrukt Leistung auch in einen entsprechenden Zusammenhang

stellen. Als Grundlage dient abermals eine Analyse der entsprechenden Stelle. Mit Hilfe einer Tätigkeitsanalyse versuchen wir, das Verhalten, das gute Leistung definiert, herauszukristallisieren. Dies kann mit Hilfe von Beobachtungen wie sie schon von Taylor (1911) unternommen wurden, mit Hilfe von Fragebögen (z. B. McCormick, Jeanneret & Mecham, 1972), Interviews (z. B. Flanagan, 1954) oder einer Kombination aus solchen Methoden (z. B. Fine & Cronsaw, 1999) geschehen. Diese Tätigkeitsanalyse dient zwei Zwecken: Zum ersten dient sie als Grundlage für die Überlegung, wie man das resultierende Kriterium „Leistung“ denn letztlich messen möchte. Zum anderen dient sie als Grundlage der Anforderungsanalyse. Aus dem für gute Leistung zu zeigenden Verhalten soll abgeleitet werden, welche psychologischen Konstrukte einer solchen Leistung zugrunde liegen. Im Allgemeinen wird hierbei Leistung verstanden als eine Funktion aus den Fähigkeiten und Fertigkeiten der Person auf der einen Seite sowie ihrer Motivation auf der anderen (Locke, Mento & Katcher, 1978). Dabei kann man jeden dieser beiden Faktoren weiter unterteilen: Motivation ist definiert als die Richtung, Ausmaß und Ausdauer der Anstrengung, die für eine Aufgabe aufgewendet wird (Campbell, 1990) und hängt selbst wiederum von den Zielen, Motiven, Interessen, Bedürfnissen und der Persönlichkeit der Person ab (Kanfer, 1992). Der Bereich der Fähigkeiten wiederum umfasst sowohl das deklarative Wissen sowie das prozedurale Wissen und die Fertigkeiten, die für diese Tätigkeit relevant sind (Campbell, 1990).

Die Analyse der Konstrukte, die der Leistung zugrunde liegen, ist praktisch relevant, wenn es um die Entwicklung von möglichen Personalauswahlmethoden geht. Angesichts der Tatsache, dass man zum Zeitpunkt der Personalauswahl im Allgemeinen nicht weiß, wie gut die Leistung ist, die ein bestimmter Kandidat in dem entsprechenden Job zeigen wird, bemüht man sich mit Hilfe der zur Personalauswahl eingesetzte Verfahren, eben jene psychologischen Konstrukte zu erfassen, von denen angenommen wird, dass sie für eine gute Leistung relevant sein werden. Das Ziel ist am Ende also die erfolgreiche Vorhersage der tatsächlichen Leistung, auch wenn diese tatsächliche Leistung am Ende kaum erfassbar ist. Anstelle dessen bemühen wir uns, möglichst genaue und vollständige Maße des Konstruktes Leistung zu entwickeln – ein ausgesprochen schwieriges Unterfangen, wie wir im Folgenden darlegen möchten.

5.2 Leistung ist multidimensional

Eine grundlegende Schwierigkeit sowohl für die Vorhersage als auch Bewertung von Leistung ist die statische sowie auch dynamische Multidimensionalität dieses Konstruktes. Anders gesagt: Zu jedem beliebigen Zeitpunkt kann der Begriff „Leistung“ inhaltlich sehr unterschiedliche Dinge bedeuten und über die Zeit hinweg ist Leistung alles andere als stabil.

Somit ist Leistung ein Konstrukt, das als Maßstab bzw. Kriterium für verschiedenste organisationale Interventionen (von Verfahren der Personalauswahl über Trainings bis hin zur Arbeitsgestaltung) zwar stets relevant, aber häufig nicht so sauber und nachhaltig zu erfassen ist, wie es eigentlich nötig wäre.

Im Folgenden gehen wir zunächst genauer auf die Multidimensionalität des Konstruktes „Leistung“ ein, bevor wir erst konzeptionell und dann konkret methodisch diskutieren, wie man dieses Konstrukt messen kann.

5.2.1 Statische Multidimensionalität von Leistung: Inhaltliche Modelle der Leistung

Zu jedem beliebigen Zeitpunkt ist Leistung multidimensional. So kann man unterscheiden, auf welche inhaltlichen Aspekte von Leistung man achten möchte. Inhaltliche Modelle der Leistung beschäftigen sich vor allem mit unterschiedlichen Facetten der Leistung. Dies können Facetten der Aufgabenleistung sein, die direkt relevant sind, um den der Organisation zur Verfügung stehenden Input in einen Output zu verwandeln – oder diesen Prozess zu unterstützen. Oder es sind Leistungsfacetten, die auf den ersten Blick nicht relevant für die Aufgabenleistung erscheinen, auf den zweiten jedoch sehr hilfreich sind, da sie wie das Öl im Getriebe das System am Laufen halten.

Ursprüngliche Modelle der Arbeitsleistung bezogen sich vor allem auf inhaltliche Kategorisierungen von Aufgabenleistung im engeren Sinn, d. h. die Leistung von Mitarbeitern auf den Facetten, die im Rahmen einer aufgabenorientierten Arbeitsplatzanalyse als notwendig identifizierbar sind und somit formal als Teil der Arbeit erkannt werden.

Laut Campbell (1990; Campbell, McCloy, Oppler & Sager, 1993) handelt es sich hierbei um drei Komponenten, die konzeptionell für jede Art von Tätigkeit relevant sind, nämlich (a) tätigkeitsspezifische Aufgabenbewältigung, (b) das Ausmaß und die Konsistenz an Mühe, die sich Mitarbeiter geben und (c) ihre persönliche Disziplin, d. h. das Ausmaß in dem man negatives Verhalten wie z. B. Absentismus, Substanzmissbrauch oder Regelverstöße vermeidet. Jedoch gäbe es je nach Tätigkeit noch weitere Komponenten der Leistung, die für einige, wenn auch nicht alle Jobs relevant wären wie nicht tätigkeitsspezifische Aufgabenbewältigung, schriftliche und mündliche Kommunikation, Führung, Management und Administration sowie die Förderung von Kollegen- und Gruppenleistung.

Gerade letzterer Punkt, die Förderung von Kollegen- und Gruppenleistung, hat eine ganze Literatur beflügelt. Organ (1988) bemerkte, dass ein Fokus auf die

reine Aufgabenleistung nur unzureichend das Verhalten abbildet, das viele Mitarbeitende tagtäglich in Organisationen sowohl zum Wohlergehen der Organisation als dem ihrer Kollegen zeigen. Dies ist der Bereich des Organizational Citizenship Behaviors (OCB), d. h. von freiwilligem Verhalten, das über die eigentlichen Aufgaben von Mitarbeitern hinausgeht und das man so i. A. auch nicht explizit von Mitarbeitern einfordern kann (Organ, 1988). Williams und Anderson (1991) zeigten, dass sich solch OCB auf zwei Ziele beziehen kann, nämlich auf die Organisation selbst (OCB-O; z. B. durch freiwillige Übernahme weiterer Aufgaben oder dem positiven Darstellen der Organisation gegenüber dritten) oder auf Individuen, die im Kontakt mit der Organisation stehen (OCB-I; z. B. indem man Kollegen hilft, Unterstützung anbietet, oder sich ausgesprochen freundlich verhält). Podsakoff, MacKenzie, Paine und Bacharach (2000) dagegen identifizierten sieben relevante Dimensionen, nämlich (a) Hilfeverhalten (Altruismus, Höflichkeit, Friedensstiftung), (b) Sportsgeist, d. h. Tolerieren der unvermeidlichen Negativitäten der Arbeit ohne zu jammern und sich zu beschweren, (c) Konformität gegenüber der Organisation, d. h. Akzeptieren und Befolgen von organisationalen Normen bezüglich dessen, was einen „guten Mitarbeiter“ ausmacht, (d) Loyalität gegenüber der Organisation, d. h. Identifikation mit der Organisation, Unterstützung und Verteidigung der organisationalen Ziele, (e) Eigeninitiative, (f) ein aktives Interesse am Leben der Organisation und eine Beteiligung an politischen Prozessen in der Organisation sowie (g) persönliche Weiterentwicklung.

Gerade der letzte Faktor wiederum spricht auch eine weitere mögliche Dimension der Arbeitsleistung an: So bemerkten Pulakos und Kollegen (Pulakos, Arad, Donovan & Plamondon, 2000), dass eine Berücksichtigung von Aufgabenleistung und OCB alleine noch stets recht statisch ist und nicht unbedingt den sich wechselnden Anforderungen der Arbeitswelt gerecht werden kann. Sie fügten dementsprechend noch eine weitere Dimension zur Arbeitsleistung hinzu, die sich gerade mit dem Umgang mit besonderen Anforderungen beschäftigt, nämlich der Anpassungsleistung von Mitarbeitern, d. h. ihrer Fähigkeit, all die oben genannten Dimensionen auch dann zur Anwendung zu bringen, wenn man sich in Krisen- oder Stresssituationen oder unter unsicheren Arbeitsbedingungen befindet. Weitere Dimensionen der Anpassungsleistung beschäftigen sich mit dem Erlernen neuer Aufgaben, Technologien und Verfahren, dem kreativen Problemlösen, zwischenmenschlichen Fertigkeiten, kultureller Aufgeschlossenheit und körperlicher Anpassungsfähigkeit.

Insgesamt sammelt sich somit unter dem Begriff „Leistung“ ein sehr weit verzweigtes Netz an Konstrukten. Hieraus ergibt sich die Frage, mit welcher Spezifität man Leistung am besten beschreiben soll. Soll man sehr differenzierte Modelle nutzen, die sich u. U. nicht auf alle Tätigkeiten generalisieren lassen oder sehr allgemeinen Modellen, die u. U. zu breit angelegte Konstrukte beinhalten

und dabei relevante Aspekte der Leistung übersehen (z. B. Scullen, Mount & Judge, 2003). Damit zusammenhängend stellt sich auch die Frage nach der wirklichen Differenzierbarkeit zwischen verschiedenen Dimensionen. So bestätigen konfirmatorische Faktorenanalysen zwar die Unterscheidung von Aufgabenleistung und OCB (Conway, 1999; Johnson, 2001), doch gleichzeitig korrelieren beide Faktoren beachtlich miteinander: So fanden Podsakoff und Kollegen (Podsakoff, Whiting, Podsakoff & Blume, 2009) korrigierte Korrelationen zwischen der Aufgabenleistung von Mitarbeitern und ihrem OCB von .47 (OCB-I) und .54 (OCB-O). Die Korrelationen von OCB mit Einschätzungen der Arbeitsleistung durch Vorgesetzte ist tendenziell sogar noch höher – ein interessanter Befund angesichts der Konzeptualisierung von OCB als Verhalten, das über die Anforderungen der Stelle hinausreicht. In diesem Sinne ist es dann auch nicht erstaunlich, dass Viswesvaran und Ones (2000) einen gemeinsamen Faktor der Arbeitsleistung identifizieren, der gute 50 % der Varianz verschiedener Leistungsindikatoren erklärt. Die Frage ist daher, wie man solche Befunde integriert.

Ein Lösungsvorschlag hierzu ist das hierarchische Modell von Bartram (2005). Dieses Modell beschreibt Leistung auf verschiedenen Abstraktionsniveaus. Arbeitsleistung setzt sich demnach aus zwei aus dem sozioanalytischen Ansatz von Hogan (e. g., Hogan & Shelton, 1998) abgeleiteten globalen Komponenten zusammen. Die beiden Komponenten stellen dar, wie sehr Verhaltensweisen dem Erreichen von Zielen dienen (Komponente 1, „getting ahead“, dies entspricht weitgehend der klassischen Aufgabenleistung) oder, wie sehr sie die Zusammenarbeit von Menschen befördern (Komponente 2, „getting along“, entsprechend dem OCB). Diese globalen Komponenten wiederum lassen sich in acht empirisch validierte Faktoren unterteilen, die „great eight“.

Zu dem Bereich des „getting ahead“ gehören hierbei sechs Faktoren:

1. *Führung und Entscheidung*, d. h. Kontrolle und Führung übernehmen, Handlungen initiieren, Richtung vorgeben und Verantwortung übernehmen.
2. *Interaktion und Präsentation*, d. h. effizientes Kommunizieren und Netzwerken, sicherer und entspannter Umgang mit anderen, andere überzeugen und beeinflussen.
3. *Analysieren und Interpretieren*, d. h. gute analytische Fertigkeiten, den Kern auch komplexer Sachverhalte erkennen, effektive Anwendung eigener Expertise.
4. *Kreieren und Konzeptualisieren*, d. h. neue Ideen und Erfahrungen gut integrieren. Suche nach Lernmöglichkeiten. Innovativität und Kreativität. Breites und strategisches Denken. Unterstützen und Vorantreiben organisationalen Wandels.

5. *Organisieren und Ausführen*, d. h. Vorausplanen und systematisches und organisiertes Arbeiten. Befolgen von Anweisungen und Prozeduren, Ergebnisse erbringen, die den abgesprochenen Standards entsprechen.
6. *Unternehmen und Leistung*, d. h. Fokus auf Ergebnisse und der Erreichung persönlicher Arbeitsziele. Zeigt sich vor allem, wenn Bemühen und Leistung eng verzahnt sind. Verständnis für geschäftliche und finanzielle Belange. Möglichkeiten der beruflichen Weiterentwicklung aufsuchen.

Zu dem Bereich des „getting along“ wiederum gehören die beiden letzten Faktoren:

7. *Unterstützen und Kooperieren*, d. h. Unterstützung anderer und Respekt in sozialen Situationen. Gute Zusammenarbeit mit Individuen und Teams, Klienten und Mitarbeitern. Konsistentes Verhalten mit klaren persönlichen Werten.
8. *Anpassung und Bewältigung*, d. h. gute Anpassung an Veränderungen und konstruktiver Umgang mit Druck und Rückschlägen.

Entsprechend Modellen der Persönlichkeit geht Bartram (2005) weiterhin davon aus, dass jeder dieser Faktoren noch weiter unterteilt werden kann, so dass er auf der nächst tieferen Hierarchieebene 20 Dimensionen darstellt, die sich dann wiederum in 112 Komponenten aufteilen. Zum Beispiel unterteilt sich der Faktor „Organisieren und Ausführen“ in drei Dimensionen wie z. B. „Planung und Organisation“, welche sich wiederum in fünf spezifischere und damit handlungsnahere Komponenten aufteilen lässt (Zielsetzung, Planung, Zeitmanagement, Ressourcenmanagement, Fortschrittsüberwachung).

5.2.2 Dynamische Multidimensionalität von Leistung

Neben der oben erwähnten inhaltlichen Vielfalt der Dimensionen der Leistung ergibt sich noch das Problem, dass die Leistung eines Mitarbeiters nicht stabil ist, sondern über Tage und Situationen hinweg fluktuiert. Während ein Teil dieser Variation eher zufällig ist, ist ein gewisser Teil jedoch auch sehr gut vorhersehbar, zum Beispiel durch die Frage, ob sich die betreffenden Mitarbeiter gerade in einer typischen oder einer maximalen Leistungssituation befinden. Unter typischen Leistungssituationen, d. h. alltäglichen Arbeitssituationen hängt die Arbeitsleistung von Mitarbeitern neben externen Zufallsfaktoren sowohl von ihren Fähigkeiten (d. h. ihrem deklarativen Wissen und ihrem prozeduralen Wissen und Fertigkeiten) und ihrer jeweiligen Motivation ab (d. h. der Richtung, dem Ausmaß, und der Ausdauer ihrer Anstrengungen). In maximalen Leistungssituationen jedoch, d. h. unter kurzzeitigen Evaluierungsbedingungen, während

derer die Instruktion, die Anstrengungen zu maximieren, offensichtlich ist, ist die Motivation der Betroffenen zwangsläufig hoch und die gezeigte Leistung eher eine Frage ihrer jeweiligen Fähigkeiten (Klehe & Anderson, 2007; Sackett, Zedeck & Fogli, 1988).

Empirisch ist der Zusammenhang zwischen Leistung in typischen und in maximalen Leistungssituationen eher gering (Beus & Whitman, 2012), und dies unabhängig davon, ob es sich um physische (Sackett et al., 1988), administrative (Klehe & Anderson, 2007) oder soziale (Klehe & Latham, 2008; Ployhart, Lim & Chan, 2001) Tätigkeiten handelt.

Weiterhin hängen die gemessene Leistung und ihre Zusammenhänge mit möglichen, etwa aus Verfahren der Personalauswahl (vgl. Abschnitt 4 in diesem Kapitel) abgeleiteten, Prädiktoren auch von zeitlichen Aspekten ab. Zeitliche oder dynamische Dimensionalität impliziert, dass Leistung sich über die Zeit hinweg verändert. Diese Veränderung kann jede von drei Formen annehmen: Erstens kann sich der Durchschnitt der Leistung verändern, z. B. weil die Leistenden als Gruppe mit der Zeit im Job effektiver (z. B. durch technische Neuerungen, die die Arbeit erleichtern) oder schlechter werden (z. B. durch physische Anforderungen, denen man mit der Zeit nicht mehr gewachsen ist), oder aber auch, weil es in Gruppen zu einem selektiven Ausscheiden von Hoch- oder Niedrig-Leistenden kommt (z. B. indem weniger erfolgreiche Mitarbeiter ihre Stelle mit der Zeit verlassen oder erfolgreiche Mitarbeiter sich einen „besseren“ Job suchen).

Zweitens kann sich die Rangordnung von Mitarbeitern bezüglich ihrer Leistung ändern, wenn einige Mitarbeiter relativ stabil in ihrer Leistung bleiben, während andere ihre Leistung über die Zeit – z. B. bedingt durch private Umstände – hinweg erhöhen oder stark verringern. Dies führt dazu, dass die meisten Leistungsprädiktoren sich besser für kurz- als für längerfristige Vorhersagen eignen (Keil & Cortina, 2001).

Damit zusammenhängend kann sich drittens auch die Validität eines jeden Leistungsprädiktors über die Zeit hinweg ändern, weil sich die Aufgaben oder die Merkmalsausprägungen der Personen über die Zeit ändern. Das „Model der Veränderung der Aufgabe“ nimmt an, dass sich z. B. aufgrund von technologischen Entwicklungen die verschiedenen Kriterien für effektive Leistung in ihrer Wichtigkeit verändern, während die individuellen relativen Fähigkeiten stabil bleiben (z. B. sinkt durch die Einführung von Scanner-Kassensystemen in Supermärkten die Anforderung an die Gedächtnisleistungen der Kassierer). Das Model der Änderung der Merkmalsausprägungen des Subjektes nimmt hingegen an, dass sich der jeweilige individuelle Level an Fähigkeiten über die Zeit verändert (z. B. aufgrund von erhöhten Erfahrungen und fachliche Kompetenzen oder verringerter physikalischer Fitness).

Entsprechend ist es möglich, dass sich die Bedeutung der Leistungsprädiktoren über die Zeit hinweg ändert. Dies zeigte sich z. B. bei Buchhaltern (Bass & Barrett, 1981). Waren anfangs vor allem Fähigkeitsmaße relevant, um Leistung vorherzusagen, so waren es später soziale Fertigkeiten (Bass & Barrett, 1981).

6 Leistungsbeurteilung

Aus der oben beschriebenen Multidimensionalität von Leistung sollte soweit schon deutlich geworden sein, dass eine vollständige und saubere Leistungserfassung alles andere als trivial und letztlich gar nicht möglich ist. Dies stellt den Kern des sogenannten *Kriteriumproblems* dar, d. h. des Problems, das das multidimensionale Konstrukt der Leistung immer nur annähernd, nie jedoch wirklich akkurat durch ein gemessenes Kriterium abgebildet werden kann. Der Grund hierfür sind zwei komplementäre Schwierigkeiten eines jeden operationalen Kriteriumsmaßes: Zum einen die *Kriteriumsdefizienz* und zum anderen die *Kriteriumskontamination*. Mit Kriteriumsdefizienz sind die Aspekte von Leistung gemeint, die durch ein operatives Kriteriumsmaß nicht abgedeckt werden können, z. B. wird kaum ein Kriteriumsmaß alle 112 Komponenten von Bartrams (2005) Modell der Arbeitsleistung vollständig abdecken können, von dynamischen Effekten ganz zu schweigen. *Kriteriumskontamination* meint weitere, nicht leistungsrelevante Aspekte, außerhalb der Kontrolle der bewerteten Person, die das operationale Kriteriumsmaß systematisch oder unsystematisch beeinflussen.

In der Praxis sollten Maße der Leistungsbeurteilung mindestens drei unabhängigen Anforderungen genügen:

1. Die gemessene Leistung muss für die Ziele der Organisation relevant sein. Auch wenn dies in der Theorie offensichtlich klingt, basiert die Kriterienauswahl in der Praxis oft weniger auf Relevanz als auf Bequemlichkeit (z. B. indem Gelegenheitsdaten (umgangssprachlich auch „eh da Daten“ genannt) genutzt werden, auch wenn sie nicht unbedingt die relevantesten Kennziffern der Leistung darstellen), auf Gewohnheit, oder auf dem Kopieren dessen, was andere Organisationen nutzen.
2. Kriteriumsmaße müssen zwischen mehr und weniger effektiven Arbeitnehmern unterscheiden. Dies benötigt (a) einen engen Bezug des Kriteriumsmaßes zu der tatsächlichen Leistung des Arbeitnehmers im Job (z. B. durch die Möglichkeit für Bewerter, die Leistung der bewerteten Person auch tatsächlich beobachten und bewerten zu können) und (b) Varianz unter den Arbeitnehmern.
3. Leistungsbeurteilungen müssen praktikabel sein, denn selbst das am besten designte Evaluationssystem wird versagen, wenn das Management mit umfangreichen Datenaufzeichnungen und -berichten konfrontiert wird, ohne einen angemessenen Ertrag für ihren Mehraufwand zu erhalten.

6.1 Arten der Leistungsbewertung

Wie wird das Kriterium „Leistung“ in Organisationen gemessen?

6.1.1 Objektive Maße der Leistungsbewertung

In einigen Bereichen sehr häufig ist die Erfassung objektiver Maße – z. T., weil diese sowieso automatisch miterfasst werden (Gelegenheitsdaten), ihre Erfassung also keinen Extra-Aufwand bedeutet, z. T. aber auch, weil diese objektiven Maße häufig recht eng mit den Zielsetzungen der jeweiligen Abteilung verbunden sind und ihre Aufgabenrelevanz daher außer Frage steht.

Klassisch und wahrscheinlich am bedeutendsten in diesem Bereich ist die Erfassung des jeweiligen quantitativen Ergebnisses, sei es in produzierten oder verkauften Stückzahlen, dem finanziellen Umsatzvolumen, der Anzahl hinzugekommener Kunden, etc. Neben dem direkten Bezug zu den Organisations- oder Abteilungszielen bietet sich die Verwendung solcher Ergebnismaße besonders dort an, wo Ergebnisse auf ganz unterschiedliche Arten erreicht werden können und/oder eine genaue Verhaltensbeobachtung zum Zwecke der Leistungsbewertung schwierig ist. Gleichzeitig führt eine reine Ergebnis-Orientierung schnell zu ungewollten Nebenwirkungen, wenn Mitarbeiter versuchen, ihre Ergebniszahlen auf Kosten der Qualität zu steigern und es so zu fehlerhaften Produkten oder kurz angebundenen Mitarbeitern der Service-Hotline kommt. Aus diesem Grund erfassen viele Systeme auch objektive Indikatoren für die Qualität der Arbeit wie z. B. die Anzahl der identifizierten Fehlproduktionen, die Anzahl der eingegangenen Beschwerden, oder den Prozentsatz an Fällen, die eine Nachbearbeitung benötigen. Was als Schwierigkeit jedoch bleibt, ist eine relative Überbewertung des Zähl- und Messbaren gegenüber weniger numerisch abbildbaren Zielsetzungen, sowie ein Mangel an Feedbackmöglichkeiten, um Mitarbeiter darin zu unterstützen, ihre Leistung zu steigern. Schließlich sind objektive Kriterien oft in einem nicht unerheblichen Maße kontaminiert.

6.1.2 Subjektive Maße der Leistungsbewertung

Weiterhin ist die Erfassung rein objektiver Maße auch deshalb problematisch, da es sich hierbei ‚nur‘ um die Ergebnisse der Mitarbeiterleistung handelt, nicht jedoch um die Leistung, d. h. das Verhalten der Mitarbeiter (Campbell et al., 1993), selbst.

Aus diesem Grund, und auch, da objektive Maße häufig nicht in der Lage sind, alle erwünschten Dimensionen von Leistung adäquat darzustellen, setzen

Organisationen häufig auch subjektive Maße der Leistungsbeurteilung ein, d. h. menschliche Bewertungen der Leistung der Mitarbeiter. Hier wiederum lässt sich unterscheiden zwischen relativen und absoluten Verfahren, wobei sich letztere noch in qualitative versus quantitative Verfahren unterteilen lassen.

6.1.2.1 Relative Bewertungssysteme

Relative Bewertungssysteme fordern die bewertende Person auf, verschiedene zu bewertende Personen relativ zueinander zu bewerten, z. B. indem Mitarbeiter in Bezug auf ihre Leistung auf einem bestimmten Kriterium in eine Rangfolge gebracht werden, jeweils paarweise miteinander verglichen, oder in einer zuvor festgelegten Verteilung (z. B. ähnlich einer Normalverteilung) einsortiert werden sollen.

Die Vorteile solch relativer Bewertungssysteme liegen zum einen in ihrer Eignung in Vorbereitung von Wahlentscheidungen (z. B. wenn es darum geht, aus den derzeitigen Mitarbeitern eine Person für eine anspruchsvollere Position auszuwählen, so liegt es nahe, die Person zu wählen, die hier im Vergleich zu anderen am besten bewertet wird) und zum anderen in der Vermeidung von bestimmten Verzerrungstendenzen. So vermeiden es Bewertende ohne solch ein relatives Bewertungssystem häufiger, zwischen ihren Mitarbeitern zu differenzieren – mit dem Ergebnis, dass alle Mitarbeiter, unabhängig von ihrer eigentlichen Leistung, am Ende durchschnittliche oder gute Leistungen attestiert bekommen.

Gleichzeitig sind solche relativen Bewertungssysteme der zukünftigen Zusammenarbeit aber nicht immer förderlich. Zum einen fordern sie sprichwörtlich zum sozialen Vergleich heraus, was zu Missgunst und Konkurrenz zwischen Mitarbeitern beitragen kann. Zum anderen sind sie i. A. nicht verhaltensspezifisch, mit dem Ergebnis, dass sich Mitarbeitende am Ende mehr mit ihrer persönlichen Position im Gesamtgefüge einer Arbeitsgruppe beschäftigen als damit, wie sie letztlich ihre Arbeit besser erledigen können. Aus diesen Gründen bevorzugen viele Organisationen absolute – und idealerweise verhaltensorientierte – Bewertungssysteme.

6.1.2.2 Absolute Bewertungssysteme

Absolute Bewertungssysteme vergleichen Mitarbeitende nicht untereinander in Bezug auf ihre Arbeitsleistung, sondern vergleichen die Leistung eines Mitarbeiters mit einem zuvor festgesetzten Standard. Man spricht diesbezüglich auch von einer kriterienorientierten Beurteilung. Weiter unterscheiden kann man hierbei zwischen qualitativen und quantitativen Verfahren.

Von einem qualitativen Verfahren spricht man i. A., wenn sich Beurteilende die Mühe machen, schriftlich festzuhalten, was ihnen am Verhalten der entsprechenden Mitarbeiter aufgefallen ist. Das kann z. B. in Form eines freien Textes passieren, der individuelle Stärken und Entwicklungsgebiete aufzeigt und Verbesserungsvorschläge unterbreitet. Der Vorteil eines solchen Verfahrens ist i. A., dass diese Vorgehensweise eine gute Vorbereitung für detailliertes Feedback darstellt. Von Nachteil ist jedoch der Mangel an Struktur und Vergleichbarkeit – und der beträchtliche Aufwand, den solch ein Verfahren mit sich bringt.

Ebenfalls qualitativ und aufwändig, jedoch für ein verhaltensorientiertes Feedback extrem hilfreich, ist ein semi-strukturierteres Festhalten von individuellen Verhalten in besonderen Leistungssituationen durch die *Critical Incident Technique* (CIT; Flanagan, 1954). Die Vorteile einer solchen Herangehensweise liegen in ihrem handlungsorientierten Arbeitsbezug. Die Aufmerksamkeit von Bewertenden und Bewerteten richtet sich dabei auf die Anforderungen der Arbeit in der jeweiligen Situation. Gerade bei weniger erfolgreich bewältigten Situationen kann man anhand des gegebenen Beispiels herausarbeiten, welches Verhalten man sich anstelle dessen von dem entsprechenden Mitarbeiter gewünscht hätte und daraus Entwicklungs- und Handlungsanleitungen für die Zukunft ableiten. Ein Problem, das auch mit der CIT jedoch bestehen bleibt, ist der extreme Aufwand für die Bewertenden und der Mangel an Quantifizierbarkeit und damit Vergleichbarkeit zwischen Mitarbeitern. Aus diesem Grunde wird die CIT in der Praxis i. A. vor allem als ergänzende Maßnahme verwendet, um quantitative Bewertungen zu begründen und um Mitarbeitern konkrete Beispiele an die Hand zu geben, anhand derer sie sich Handlungsalternativen vorstellen können.

Die eigentliche Leistungsbewertung selbst erfolgt jedoch letztlich häufig anhand von quantitativen Verfahren. Quantitativ bedeutet hierbei eigentlich nur, dass jeder Mitarbeiter anhand verschiedener Leistungsdimensionen bewertet wird. Die Vielfalt der hier möglichen Bewertungsmethoden ist jedoch groß, mit sehr unterschiedlichen konzeptionellen und methodischen Hintergründen. So kann eine jede Bewertung auf einer grafischen Skala erfolgen und/oder durch bestimmte numerische (z. B. 1–5) und/oder verbale Anker (z. B. herausragend, überdurchschnittlich, durchschnittlich, unterdurchschnittlich) begleitet werden. Doch wie schaut es mit den Leistungsdimensionen aus, die zu bewerten sind? Letztlich befindet man sich hier in einem Dilemma: Einerseits verlangen praktische Überlegungen (z. B. die Tatsache, dass Vorgesetzte häufig neben dem üblichen Tagesgeschäft eine ganze Reihe von Mitarbeitern parallel zu beurteilen haben) möglichst knapp gefasste und daher schnell auszufüllende Beurteilungsmethoden. Andererseits jedoch tragen solch knapp gefassten und daher grobkategorialen Bewertungsverfahren die Gefahr in sich, als Eigenschaftszuschrei-

bungen interpretiert zu werden – und damit wenig hilfreich zu sein in dem Anliegen, Leistung nicht nur zu bewerten, sondern durch ein entsprechendes Feedback auch verbessern zu können.

Im Folgenden beschreiben wir hier zwei quantitative Bewertungssysteme, die sich in der Praxis besonders bewährt haben und beide den wissenschaftlichen Anforderungen an Reliabilität und Validität vollauf genügen. Die Grundlage dieser Verfahren bietet dabei wieder die CIT (Flanagan, 1954), wenn dieses Mal auch in Funktion einer Tätigkeitsanalyse. D. h. hierbei dient die CIT nicht dazu, das Verhalten eines bestimmten Mitarbeiters in kritischen Situationen zu beschreiben, sondern relevante Verhaltensweisen für die entsprechende Tätigkeit zu identifizieren.

Behavioral Observation Scales (BOS, Latham & Wexley, 1977) sammeln die durch die CIT identifizierten Verhaltensweisen, die für eine Position wünschenswert wären, einfach in einer Liste und bitten dann die bewertende Person, für jede dieser Verhaltensweisen einzuschätzen, wie häufig die zu bewertende Person diese Verhaltensweisen im Allgemeinen an den Tag legt (z. B. auf einer Likert-Skala von 1 = *fast nie* bis 5 = *fast immer*) (vgl. Abb. 4).

		fast nie			fast immer	
1	Beschreibt Mitarbeitern die Einzelheiten der Veränderung	1	2	3	4	5
2	Erklärt, warum die Veränderung notwendig ist	1	2	3	4	5
3	Bespricht, inwiefern die Veränderung die Mitarbeiter betrifft	1	2	3	4	5
4	Hört sich die Bedenken der Mitarbeiter an	1	2	3	4	5
5	Bittet die Mitarbeiter um Unterstützung bei der Umsetzung der Veränderung	1	2	3	4	5
6	Setzt, falls notwendig, einen Besprechungstermin wegen der Bedenken der Mitarbeiter fest	1	2	3	4	5
Auswertung:						
Nicht angemessen		Angemessen		Gut		Sehr gut
6–10		11–15		16–20		21–25

Abbildung 4:

Beispiel einer Behavioral Observation Scale (BOS) – Widerstand gegenüber Veränderung überwinden (aus Latham & Wexley, 1981).

Vorteile dieses Verfahrens sind, dass es einerseits sehr leicht zu entwickeln und andererseits für beurteilende Personen sehr leicht auszufüllen ist, da man sich

pro Frage auf jeweils eine relevante Verhaltensweise bezieht. Diese eindeutige Verhaltensorientierung gibt weiterhin für den bewerteten Feedbackempfänger klare Hinweise darauf, welche Verhaltensweisen für die Zukunft wünschenswert wären. Neben diesen Vorteilen möchten wir zwei Nachteile der BOS herausstellen. Auch hier ist der beträchtliche Aufwand, den eine Bewertung auf verschiedenen Leistungsdimensionen – und wahrscheinlich auch über verschiedene Leistungsträger – für die bewertende Person mit sich bringt, beträchtlich. Darüber hinaus besteht die Gefahr, dass Bewertende anhand der konkret erfassten Verhaltensweisen nicht unbedingt auf die allgemeine Leistungsdimension generalisieren und somit Verhalten, das zwar relevant ist, aber nicht konkret erfragt wird, eventuell aus der Bewertung herausfällt. Dass relevante Verhaltensweisen nicht in die Skala aufgenommen werden kann beispielsweise daran liegen, dass dieses Verhalten, wenn es vorkommt, zwar desaströse Auswirkungen hat, aber eben so selten vorkommt, dass es für die meisten bewerteten Personen nicht relevant ist.

Diese Nachteile der BOS sind behoben im Format der Behavioral Anchored Rating Scales (BARS, Smith & Kendall, 1963). Hier werden zu einer konkreten Situation sowohl positive als auch negative Verhaltensbeispiele aus der CIT gesammelt und dann zu Dimensionen zusammengefasst. Während die Beurteiler bei BOS in der Regel Häufigkeitsurteile treffen (z. B. wie häufig wurde eine erfolgskritische positive Verhaltensweise gezeigt), beurteilen Sie bei BARS, ob das Verhalten der Mitarbeiter eher dem positiven oder negativen (oder den Zwischenstufen) Anker-Verhalten der Skala entspricht.

Das Ergebnis ist i. A. ein mit wenig Zeitaufwand handhabbares Bewertungssystem, das in seinem Kern verhaltensbasiert ist. Jedoch ist die Entwicklung dieses Verfahrens im Vergleich zu den BOS zeitaufwändiger. Smith und Kendall (ebd.) gehen hier bei der Entwicklung der Skalen von der Zusammenarbeit von vier unabhängigen Gruppen von Experten (z. B. Stelleninhaber, Vorgesetzte und Personalverantwortliche) für die jeweilige Position aus. In der Praxis wird allerdings häufig aus praktischen Gründen mit weniger Gruppen gearbeitet, die dann in verschiedenen Stadien der Entwicklung wiederholt zum Einsatz kommen: Die erste Gruppe definiert und beschreibt die relevanten Leistungsdimensionen, die zweite sammelt konkrete Verhaltensbeispiele für jeweils gute, mittlere oder schwache Leistung auf dieser Dimension. Die dritte Gruppe erhält Dimensionsbeschreibungen und Verhaltensbeispiele getrennt voneinander vorgelegt und muss die Verhaltensbeispiele jeweils der korrekten Dimension zuordnen, wodurch Verhaltensbeispiele unklarer Zugehörigkeit gestrichen werden können. Mitglieder der vierten Gruppe schließlich geben subjektive Urteile für jede der Verhaltensweisen ab, welchem Leistungsniveau (z. B. von 1 – *extrem schwach* bis 9 – *herausragend*) auf der entsprechenden Dimension diese Verhaltensweisen jeweils entsprechen.

Nur Verhaltensweisen, bei denen sich Experten in ihrer Bewertung halbwegs einig sind, d. h. bei denen ihre Bewertung eine geringe Standardabweichung aufweist, werden schließlich in die BARS mit aufgenommen und hier entsprechend ihrer jeweiligen Mittelwerte über die Skala hin verteilt mit dem Ziel, am Ende repräsentative Beispiele über den gesamten Bereich des zu erwartenden Leistungsspektrums hinweg zu erhalten.

Am Ende bildet eine sauber entwickelte BARS damit eine solide Grundlage für eine effiziente und valide Leistungsbeurteilung. Schwierigkeiten kann es jedoch u. U. geben, wenn ein und dieselbe zu bewertende Person Verhaltensweisen zeigt, die sowohl den Anker für herausragende als auch für schwache Leistungen entsprechen und ein übergreifendes Gesamturteil ihr damit nicht unbedingt gerecht wird. Auch zeigen BARS in ihrem aggregierten Format zwar an, welche Verhaltensweisen mehr oder weniger wünschenswert sind, geben jedoch nicht wirklich eindeutig an, welche Verhaltensweisen für eine konkrete Bewertung letztendlich maßgeblich waren. Der Nutzen für Feedback und Mitarbeiterentwicklung ist daher u. U. etwas geringer, als bei dem oben beschriebenen BOS-Format.

Eine Anmerkung zum Ende: Wie schon die CIT ist das Format der BARS nicht nur hilfreich im Bereich der Leistungsbeurteilungen, sondern auch für die Tätigkeitsanalyse (z. B. Fleishmans Job Analysis Survey nutzt Items im BARS-Format, um die für eine Stelle nötigen Kompetenzen einschätzen zu lassen) und die Personalauswahl (z. B. entsprechen die Antwortanker strukturierter Interviews letztlich recht grob-gehaltenen BARS).

6.1.3 Subjektive Maße der Leistungsbewertung: Wer bewertet?

Neben den verschiedenen Formaten der subjektiven Leistungsbewertung stellt sich natürlich auch die Frage, wer qualifiziert und interessiert daran ist, die Leistung von Mitarbeitern einzuschätzen. Die Leistungsbeurteilung ist eine der Führungstätigkeiten, die dem direkten Vorgesetzten vorbehalten ist. Das hat den Vorteil, dass dies i. A. die Person ist, die administrative Entscheidungen über diesen Mitarbeiter zu treffen hat und im Zweifelsfalle für einen Mangel an Leistung verantwortlich gemacht wird. Auch können Vorgesetzte i. A. besser als andere Beurteiler die Arbeitsleistung in Relation zu den eigentlichen Zielen der jeweiligen Organisationseinheit setzen. Gleichzeitig jedoch sind Vorgesetzte nicht jederzeit vor Ort und können daher auch nicht immer das Verhalten ihrer Mitarbeiter repräsentativ beobachten. Es gibt auch Überlegungen, ob nicht die Anwesenheit der Vorgesetzten selbst das Verhalten beeinflusst, dass Mitarbeiter zu zeigen bereit sind. So fanden Sackett und Kollegen (1988), dass Vorgesetzten-

urteile über Leistungen eher mit der objektiven Leistung in maximalen Leistungssituationen als in typischen Situationen zusammenhängen. Sackett und Kollegen (ebd.) erklärten diesen Effekt zwar mit besonderen Eigenheiten der in ihrer Studie untersuchten Stellen, doch ist es angesichts der Macht, die Vorgesetzte über ihre Mitarbeiter haben, gut möglich, dass sich dieses Ergebnis – zumindest unter bestimmten Umständen – auch auf andere Stellen generalisieren lässt (vgl. Klehe & Anderson, 2007; Thorsteinson & Balzer, 1999).

Als mögliches Pendant, das auch die typische Leistung von Mitarbeitern Tag ein Tag aus zu sehen kriegt und häufig auch direkt davon betroffen ist, können Kollegen als Quelle der Leistungsbeurteilung dienen. So hat sich auch gezeigt, dass Kollegen i. A. recht reliabel und valide urteilen (Harris & Schaubroek, 1988). Gleichzeitig jedoch sind auch Kollegen nicht immer unparteiisch, sondern u. U. bei ihren Bewertungen durch Ähnlichkeiten, Freundschaften oder die Konkurrenz um Ressourcen beeinflusst. Ein System von Leistungsbeurteilungen durch Kollegen kann dazu führen, dass die Zufriedenheit und der Zusammenhalt zwischen den Mitarbeitern leidet, was sich dann summa summarum auch negativ auf die Arbeitsleistung auswirken wird (Beal, Cohen, Burke & McLendon, 2003; Judge, Thoresen, Bono & Patton, 2001).

Daneben gibt es, je nach Tätigkeit, noch die Möglichkeit, eigene untergebene Mitarbeiter und Kunden um eine Bewertung zu bitten. Gerade die Nutzung von Mitarbeiterurteilen kann in verschiedener Hinsicht sehr hilfreich sein. So erleben Mitarbeiter ihre Vorgesetzte häufig unter verschiedenen Bedingungen. Auch sind sie häufig bei suboptimalem Führungsverhalten die primär Leidtragenden. Während persönliche Rückmeldungen von Mitarbeitern gegenüber ihren Vorgesetzten im Alltag angesichts des Machtgefälles zwischen beiden Seiten häufig eher zu kurz kommen, bietet sich Mitarbeitern durch eine strukturierte Leistungsbeurteilung ihres Vorgesetzten eine Möglichkeit, Gehör zu finden und somit auch eventuelle Änderungen anzustoßen. Wichtig hierbei ist jedoch gerade angesichts des Machtungleichgewichtes zwischen Vorgesetzten und Mitarbeitern einer genügenden Anonymität der Bewertungen. Es muss gewährleistet sein, dass sich die Mitarbeiter keine Sorgen machen müssen, ob ihre Bewertungen auf sie persönlich zurückzuführen sind. In der Praxis wird dieses Verfahren daher i. A. erst eingesetzt, wenn die betreffende Arbeitsgruppe genügend groß ist und auch genügend Personen (meist mindestens drei, lieber mindestens fünf) an den jeweiligen Bewertungen teilnehmen.

Die Nutzung von Kunden als Quelle der Leistungsbeurteilung ist sicherlich sinnvoll, wenn der Kundenkontakt eine der kennzeichnenden Charakteristika der Stelle ist. Gleichzeitig können Kunden i. A. die Ziele der Organisation nicht gut beurteilen und haben unter Umständen selbst sogar Ziele, die denen der Organisation diametral entgegenstehen (z. B. wenn es darum geht, welche Leis-

tungen durch den betreffenden Mitarbeiter in Rechnung gestellt werden, wo Kulanz gezeigt wird, etc.).

Schließlich gibt es noch die zu bewertende Person selbst, die gebeten werden kann, sich zu bewerten. Ein Vorteil hierbei ist, dass die betroffene Person selbst i. A. viel besser nachvollziehen kann, warum sie sich in verschiedenen Situationen so verhalten hat, wie sie es getan hat. Externe Beurteiler laufen demgegenüber eher Gefahr, dem fundamentalen Attributionsfehler anheim zu fallen und die Situation zu vernachlässigen. Gleichzeitig tendieren Menschen dazu, sich im Sinne eines „self-serving biases“ recht milde und positiv verzerrt einzuschätzen (Harris & Schaubroeck, 1988). Bedauerlicher Weise ist diese Selbstüberschätzung unter tatsächlich schwächer leistenden Mitarbeitern noch stärker ausgeprägt als unter Leistungsträgern. Als Quelle, um die tatsächliche Leistung von Mitarbeitern einzuschätzen, taugen Selbstbeurteilungen daher i. A. reichlich wenig, was auch der Grund sein kann, warum Selbstbeurteilungen wesentlich weniger mit anderen Quellen der Leistungsbeurteilung korrelieren als es zwischen den übrigen diskutierten Quellen der Fall ist (Conway & Huffcutt, 1997). Gleichzeitig jedoch sind auch Selbstbeurteilungen sehr sinnvoll, da sie i. A. die Motivation der Mitarbeiter erhöhen, sich mit ihrem Leistungsprofil und den darin enthaltenen individuellen Stärken und Schwächen auseinanderzusetzen. Damit erhöht sich auch ihre Offenheit für das Feedback durch andere.

Die genannten Quellen (Selbst, Vorgesetzte, Kollegen, Mitarbeiter und Kunden) können zu einem multiperspektivischen Feedback (auch 360 Grad Feedback genannt) kombiniert werden (vgl. z. B. Scherm, 2005).

Abschließend muss noch festgehalten werden, dass all die hier behandelten Quellen der Leistungsbeurteilung bisher nur die Leistung individueller Mitarbeiter betrifft. Eine weitere Möglichkeit besteht darin, nicht die Leistung von individuellen Mitarbeitern, sondern vom Team insgesamt zu beurteilen. Dies kann sinnvoll sein, wenn der Fokus und die Ziele auf Teamebene festgelegt sind und individuelle Leistungsbeiträge schwer identifizierbar sind. In bestimmten Situationen könnte eine Bewertung individueller Leistungen sich sogar negativ auswirken, indem sie den Teamfokus reduziert.

7 Eine integrative Perspektive – Ausblick

Mit dem vorliegenden Beitrag möchten wir einen Beitrag zur Integration der Themen Personalauswahl und -beurteilung leisten. Eines der gemeinsamen Elemente beider Themen ist die Anforderungsanalyse. Auf der Basis der Anforderungsanalyse werden die Kriterien bestimmt, die gleichermaßen zur Beurteilung von Mitarbeitern und zur Beurteilung der Aussagekraft von Personalauswahl-

verfahren herangezogen werden. Bezüglich der anforderungsanalytischen Verfahren besteht ein erheblicher Forschungsbedarf zur Validität der aus Anforderungsanalysen abgeleiteten Aussagen.

Einen weiteren Themenschwerpunkt der Zukunft sehen wir in der stärkeren Berücksichtigung der Mehrdimensionalität und Dynamik der Arbeitsleistung. Sowohl die Arbeitsleistung als Kriterium als auch die Personalauswahlverfahren als Prädiktoren sollten als mehrdimensional begriffen, beforscht und gehandhabt werden. In Bezug auf die Personalauswahlverfahren bedeutet dies vor allem zu realisieren, dass Messergebnisse immer eine Einheit aus Konstrukt und Methoden sind. Erkenntnisfortschritte sind in dem Maße zu erwarten, indem man unbestimmte Begriffe wie „Assessment-Center“ in definierte Elemente von diagnostisch ähnlichen Verfahren einerseits und Konstrukten andererseits differenziert. Klassifikationssysteme für Verfahren sowie Theorien für Konstrukte bieten hier einen Orientierungsrahmen. Auf Seiten der Personalbeurteilung erscheint insbesondere eine weitere Beschäftigung mit Theorien des Berufserfolgs (einschl. der Zufriedenheit und Gesundheit) fruchtbar. Eine entsprechende Theorie sollte der Multidetermination beruflicher Leistung ebenso Rechnung tragen wie der Dynamik.

Die Praxis ist gut beraten, wenn sie sowohl zur Eignungsbeurteilung von (internen und externen) Kandidaten als auch zur Mitarbeiterbeurteilung eine gezielte Kombination von Verfahren anwendet, die den verschiedenen Dimensionen der Arbeitsleistung sowie deren Veränderung über die Zeit Rechnung trägt.

Literatur

- Ackerman, P. L. & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245. <http://doi.org/10.1037/0033-2909.121.2.219>
- American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME) (2014). *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.
- Anderson, N. R., Salgado, J. F. & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, *18*, 291–304. <http://doi.org/10.1111/j.1468-2389.2010.00512.x>
- Arthur, W., Day, E. A., McNelly, T. L. & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*, 125–154. <http://doi.org/10.1111/j.1744-6570.2003.tb00146.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*, 150–166. <http://doi.org/10.1177/1088868306294907>

- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., di Blas, L. et al. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*, 356–366.
- Back, M. D., Schmukle, S. C. & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*, 533–548. <http://doi.org/10.1037/a0016229>
- Baron-Boldt, J., Funke, U. & Schuler, H. (1989). Prognostische Validität von Schulnoten. Eine Metaanalyse der Prognose des Studien- und Ausbildungserfolgs. In R. S. Jäger, R. Horn & K. Ingenkamp (Hrsg.), *Tests und Trends 7* (S. 11–39). Weinheim: Beltz.
- Barrick, M. R., Mount, M. K. & Gupta, R. (2003). Meta-Analysis of the Relationship between the five factor model of personality and Holland's Occupational Types. *Personnel Psychology*, *56*, 45–74. <http://doi.org/10.1111/j.1744-6570.2003.tb00143.x>
- Barrick, M. R., Mount, M. K. & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9–30.
- Barrick, M. R., Shaffer, J. A. & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology*, *94*, 1394–1411. <http://doi.org/10.1037/a0016532>
- Bartram, D. (2005). The Great Eight competencies: a criterion-centric approach to validation. *Journal of Applied Psychology*, *90*, 1185–1203. <http://doi.org/10.1037/0021-9010.90.6.1185>
- Bass, B. M. & Barrett, G. V. (1981). *People, work, and organizations: An introduction to industrial and organizational psychology* (2nd ed.). Boston: Allyn and Bacon.
- Beal, D. J., Cohen, R. R., Burke, M. J. & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, *88*, 989–1004. <http://doi.org/10.1037/0021-9010.88.6.989>
- Becker, N., Höft, S., Holzenkamp, M. & Spinath, F. M. (2011). The predictive validity of assessment centers in German-speaking regions: A meta-analysis. *Journal of Personnel Psychology*, *10*, 61–69. <http://doi.org/10.1027/1866-5888/a000031>
- Beermann, D. & Kersting, M. (in Druck). People Analytics – Wie die Personalarbeit im Allgemeinen und die Eignungsdiagnostik von technischen Innovationen profitieren können. *Wirtschaftspsychologie*.
- Beus, J. M. & Whitman, D. S. (2012). The relationship between typical and maximum performance: A meta-analytic examination. *Human Performance*, *25*, 355–376. <http://doi.org/10.1080/08959285.2012.721831>
- Binning, J. F. & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*, 478–494. <http://doi.org/10.1037/0021-9010.74.3.478>
- Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational and Organizational Psychology*, *69*, 107–120. <http://doi.org/10.1111/j.2044-8325.1996.tb00603.x>
- Bowler, M. C. & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, *91*, 1114–1124. <http://doi.org/10.1037/0021-9010.91.5.1114>

- Bowling, N. A. & Burns, G. N. (2010). A comparison of work-specific and general personality measures as predictors of work and non-work criteria. *Personality and Individual Differences*, 49, 95–101. <http://doi.org/10.1016/j.paid.2010.03.009>
- Brannick, M. T., Levine, E. L. & Morgeson, F. P. (2007). *Job and work analysis: Methods, research, and applications for human resource management* (2nd ed.). Thousand Oaks: Sage Publications.
- Brunstein, J. C. (2003). Implizite Motive und motivationale Selbstbilder: Zwei Prädiktoren mit unterschiedlichen Gültigkeitsbereichen. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.), *Diagnostik von Motivation und Selbstkonzept* (S. 59–88). Göttingen: Hogrefe.
- Brunstein, J. C. & Maier, G. W. (2005). Implicit and self-attributed motives to achieve: two separate but interacting needs. *Journal of Personality and Social Psychology*, 89, 205–222. <http://doi.org/10.1037/0022-3514.89.2.205>
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (Vol. 1, 2nd ed., pp. 687–732). Palo Alto, CA: Consulting Psychologists Press.
- Campbell, J. P., McCloy, R. A., Oppler, S. H. & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco, CA: Jossey Bass.
- Campion, M. A., Fink, A. A., Ruggeberg, B. J., Carr, L., Phillips, G. M. & Odman, R. B. (2011). Doing competencies well: Best practices in competency modeling. *Personnel Psychology*, 64, 225–262. <http://doi.org/10.1111/j.1744-6570.2010.01207.x>
- Cascio, W. F. & Aguinis, H. (2011). *Applied Psychology in Human Resource Management* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Chapman, D. S. & Zweig, D. I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology*, 58, 673–702. <http://doi.org/10.1111/j.1744-6570.2005.00516.x>
- Collins, C. J., Hanges, P. J. & Locke, E. A. (2004). The relationship of achievement motivation to entrepreneurial behaviour: A meta-analysis. *Human Performance*, 17, 95–117. http://doi.org/10.1207/S15327043HUP1701_5
- Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology*, 84, 3–13. <http://doi.org/10.1037/0021-9010.84.1.3>
- Conway, J. M. & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10, 331–360. http://doi.org/10.1207/s15327043hup1004_2
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 443–507). Washington DC: American Council on Education.
- Diagnostik- und Testkuratorium (2017). *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430*. Heidelberg: Springer.
- DIN (2016). *DIN 33430: Anforderungen an berufsbezogene Eignungsdiagnostik*. Berlin: Beuth.
- Dipboye, R. L. (1989). Threats to the incremental validity of interviewer judgments. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview* (pp. 45–60). Newbury Park, CA: Sage.
- Dipboye, R. L. (1992). *Selection interviews: Process perspectives*. Cincinnati, OH: South-Western.

- Dipboye, R. L., Wooten, K. & Halverson, S. K. (2004). Behavioral and situational interviews. In J. C. Thomas (Ed.), *Comprehensive Handbook of Psychological Assessment. Volume 4: Industrial and organizational assessment* (pp. 297–316). Hoboken, NJ: John Wiley & Sons.
- 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*, 40–57.
- Duval, S. (2005). The „Trim and Fill“ method. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 127–144). Chichester, England: Wiley.
- Eagly, A. H. & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573–598. <http://doi.org/10.1037/0033-295X.109.3.573>
- Elias, P. & Birch, M. (1994). *Establishment of community-wide occupational statistics. ISCO 88. A guide for users*. Warwick: University of Warwick.
- Erpenbeck, J., Rosenstiel, L. von, Grote, S. & Sauter, W. (2017). *Handbuch Kompetenzmessung: erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (3., überarbeitete und erweiterte Auflage). Stuttgart: Schäffer-Poeschel.
- Fennekels, G. P. (1995). *PC-OFFICE 1.0. Postkorb zur Diagnose von Führungsverhalten (Handanweisung)*. Göttingen: Hogrefe.
- Fine, S. A. & Cronshaw, S. F. (1999). *Functional Job Analysis: A Foundation for Human Resources Management*. Mahwah, NJ: Lawrence Erlbaum.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327–358. <http://doi.org/10.1037/h0061470>
- Funke, J. (1993). Computergestützte Arbeitsproben: Begriffserklärung, Beispiele sowie Entwicklungspotentiale. *Zeitschrift für Arbeits- und Organisationspsychologie, 37*, 119–129.
- Geis, A. (2006). *SDM (Survey Design and Methodology). Handbuch für die Berufsvercodung (Stand: März 2011)*. Mannheim: GESIS.
- Gibbons, A. M., Rupp, D. E., Snyder, L. A., Holub, A. S. & Woo, S. E. (2006). A preliminary investigation of developable dimensions. *The Psychologist-Manager Journal, 9*, 99–123. http://doi.org/10.1207/s15503461tpmj0902_4
- Goertz, W., Hülshager, U. R. & Maier, G. W. (2014). The validity of specific cognitive abilities for the prediction of training success in Germany: A meta-analysis. *Journal of Personnel Psychology, 13*, 123–133. <http://doi.org/10.1027/1866-5888/a000110>
- Graumann, C. F. (1969). *Einführung in die Psychologie. Bd. 1: Motivation*. Bern: Huber.
- Harris, M. M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43–62. <http://doi.org/10.1111/j.1744-6570.1988.tb00631.x>
- Harvey-Cook, J. E. & Taffler, R. J. (2000). Biodata in professional entry-level selection: Statistical scoring of common format applications. *Journal of Occupational and Organizational Psychology, 73*, 103–118. <http://doi.org/10.1348/096317900166903>
- Hausknecht, J. P., Day, D. V. & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639–683. <http://doi.org/10.1111/j.1744-6570.2004.00003.x>

- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T. & Moriarty Gerrard, M. O. (2007). Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, *92*, 373–385. <http://doi.org/10.1037/0021-9010.92.2.373>
- Hell, B., Schuler, H., Boramir, I. & Schaar, H. (2006). Verwendung und Einschätzung von Verfahren der internen Personalauswahl und Personalentwicklung im 10 Jahres-Vergleich. *Zeitschrift für Personalforschung*, *20*, 58–78.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*, 47–77. <http://doi.org/10.3102/00346543058001047>
- Hoffman, B. J., Kennedy, C., LoPilato, A., Monahan, E. & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, *100*, 1143–1168. <http://doi.org/10.1037/a0038707>
- Höft, S. & Kersting, M. (2017). Anforderungsprofil, Verhaltensbeobachtung und Verhaltensbeurteilung. In Diagnostik- und Testkuratorium (Hrsg.), *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 27–63). Heidelberg: Springer.
- Höft, S., Obermann, C. & Becker, J.-N. (2014). Gibt es Besonderheiten bei kompetenzorientierten Assessment-Center-Verfahren? Eine Analyse des Praxiseinsatzes basierend auf der Anwenderbefragung des Arbeitskreises Assessment Center e. V. von 2012. *Wirtschaftspsychologie*, *16*, 95–103.
- Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, *88*, 100–112. <http://doi.org/10.1037/0021-9010.88.1.100>
- Hogan, R. & Shelton, D. (1998). A Socioanalytic Perspective on Job Performance. *Human Performance*, *11*, 129–144. <http://doi.org/10.1080/08959285.1998.9668028>
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources.
- Huffcutt, A. I., Conway, J. M., Roth, P. L. & Klehe, U. C. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment*, *12*, 262–273. http://doi.org/10.1111/j.0965-075X.2004.280_1.x
- Huffcutt, A. I., Conway, J. M., Roth, P. L. & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, *86*, 897–913. <http://doi.org/10.1037/0021-9010.86.5.897>
- Huffcutt, A. I., Culbertson, S. S. & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment*, *22*, 297–309. <http://doi.org/10.1111/ijsa.12078>
- Hülshager, U. R., Maier, G. W. & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, *15*, 3–18. <http://doi.org/10.1111/j.1468-2389.2007.00363.x>
- International Test Commission (ITC) (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, *6*, 143–171. http://doi.org/10.1207/s15327574ijt0602_4
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, *35*, 21–35.

- Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology, 86*, 984–996. <http://doi.org/10.1037/0021-9010.86.5.984>
- 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*, 857–862. <http://doi.org/10.1016/j.paid.2011.01.011>
- Joseph, D. L. & Newman, D. A. (2010). Emotional Intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology, 95*, 54–78. <http://doi.org/10.1037/a0017286>
- Judge, T. A., Bono, J. E., Ilies, R. & Gerhardt, M. W. (2002). Personality and Leadership: A Qualitative and Quantitative Review. *Journal of Applied Psychology, 87*, 765–780. <http://doi.org/10.1037/0021-9010.87.4.765>
- Judge, T. A., Heller, D. & Mount, M. K. (2002). Five-Factor Model of Personality and Job Satisfaction: A Meta-Analysis. *Journal of Applied Psychology, 87*, 530–541. <http://doi.org/10.1037/0021-9010.87.3.530>
- Judge, T. A., Higgins, C. A., Thoresen, C. J. & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*, 621–652. <http://doi.org/10.1111/j.1744-6570.1999.tb00174.x>
- Judge, T. A. & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*, 797–807. <http://doi.org/10.1037/0021-9010.87.4.797>
- Judge, T. A., Thoresen, C. J., Bono, J. E. & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127*, 376–407. <http://doi.org/10.1037/0033-2909.127.3.376>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73. <http://doi.org/10.1111/jedm.12000>
- Kanfer, R. (1992). Work motivation: New directions in theory and research. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (Vol. 7, pp. 1–53). New York: Wiley.
- Keil, C. T. & Cortina, J. M. (2001). Degradation of validity over time: A test and extension of Ackerman's model. *Psychological Bulletin, 127*, 673–697. <http://doi.org/10.1037/0033-2909.127.5.673>
- Kepes, S., Banks, G. C., McDaniel, M. A. & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15*, 624–662. <http://doi.org/10.1177/1094428112452760>
- Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie, 33*, 420–433.
- Kersting, M. (2011). Managementdiagnostik: Verfahren und Qualitätsaspekte. In C. Niedereichholz, J. Niedereichholz & J. Staudé (Hrsg.), *Handbuch der Unternehmensberatung. Organisationen führen und verwalten* (Kz. 3960, S. 1–18). Berlin: Erich Schmidt.
- Kersting, M. & Birk, M. (2011). Zur zweifelhaften Validität und Nützlichkeit von Anforderungsanalysen für die Interpretation eignungsdiagnostischer Daten. In P. Gelléri u. C. Winter (Hrsg.), *Potentiale der Personalpsychologie: Einfluss personaldiagnostischer Maßnahmen auf den Berufs- u. Unternehmenserfolg* (S. 83–95). Göttingen: Hogrefe.

- Kersting, M. & Ott, M. (2016). Diversity-gerechte Personalauswahl. Wie man die Personalauswahl gestalten muss, um Potenziale in allen Gesellschaftsgruppen zu erkennen. In T. Ringeisen & P. Genkova (Hrsg.), *Handbuch Diversity Kompetenz. Bd. 1. Perspektiven und Anwendungsfelder* (S. 679–692). Wiesbaden: Springer.
- Klehe, U.-C. & Anderson, N. R. (2007). Working hard and working smart: Motivation and ability during typical and maximum performance. *Journal of Applied Psychology*, *92*, 978–992. <http://doi.org/10.1037/0021-9010.92.4.978>
- Klehe, U.-C. & Latham, G. (2005). The predictive and incremental validity of the situational and patterned behavior description interviews for teamplaying behavior. *International Journal of Selection and Assessment*, *13*, 108–115. <http://doi.org/10.1111/j.0965-075X.2005.00305.x>
- Klehe, U.-C. & Latham, G. (2008). Predicting typical and maximum performance with measures of motivation and abilities. *Psychologica Belgica*, *48*, 67–91. <http://doi.org/10.5334/pb-48-2-3-67>
- Kolk, N. J., Born, M. P. & van der Flier, H. (2004). A triadic approach to the construct validity of the assessment center: The effect of categorizing dimensions into a feeling, thinking and power taxonomy. *European Journal of Psychological Assessment*, *20*, 149–156.
- Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: Vertiefende und weiterführende Metaanalysen. *Psychologische Rundschau*, *60*, 82–98. <http://doi.org/10.1026/0033-3042.60.2.82>
- Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*, 1–49. <http://doi.org/10.1111/j.1744-6570.1996.tb01790.x>
- Kristof-Brown, A. L., Zimmerman, R. D. & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, *58*, 281–342. <http://doi.org/10.1111/j.1744-6570.2005.00672.x>
- Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H. & Hertel, G. (2015). How „situational“ is judgment in situational judgment tests? *Journal of Applied Psychology*, *100*, 399–416.
- Kurz, R. & Bartram, D. (2001). *Competency and individual performance: Modelling the world of work. Internal SHL Memorandum*. Thames Ditton: SHL.
- Lang, J. W. B., Kersting, M., Hülshager, U. R. & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology*, *63*, 595–640. <http://doi.org/10.1111/j.1744-6570.2010.01182.x>
- Latham, G. P. & Wexley, K. N. (1977). Behavioral observation scales for performance appraisal purposes. *Personnel Psychology*, *30*, 255–268. <http://doi.org/10.1111/j.1744-6570.1977.tb02092.x>
- Latham, G. P. & Wexley, K. N. (1981). *Increasing productivity through performance appraisal. Addison Wesley series on managing human resources*. Reading, MA.: Addison-Wesley.
- Levine, S. P. & Feldman, R. S. (2002). Women and men's nonverbal behavior and self-monitoring in a job interview setting. *Applied H. R. M. Research*, *7*, 1–14.
- Lienert, G. A. (1958). *Mechanisch-Technischer Verständnistest (MTVT)*. Göttingen: Hogrefe.

- Lienert, G.A. (1967). *Drahtbiegeprobe (DBP)* (2. Aufl.). Göttingen: Hogrefe.
- Littau, P. & Hattendorf, M. (2013). *Was Personaler über Personalentwicklung denken. Eine Studie über aktuelle Herausforderungen im Umgang mit Grundbegriffen der Personalentwicklung*. Warburg: Strametz & Associates. Verfügbar unter <http://www.strametz.de/index.php/studien.html>
- Locke, E.A., Mento, A.J. & Katcher, B.L. (1978). The interaction of ability and motivation in performance: An exploration of the meaning of moderators. *Personnel Psychology*, *31*, 269–280. <http://doi.org/10.1111/j.1744-6570.1978.tb00446.x>
- Marcus, B., Höft, S. & Riediger, M. (2006). Integrity tests and the five-factor model of personality: A review and empirical test of two alternative positions. *International Journal of Selection and Assessment*, *14*, 113–130. <http://doi.org/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*, 1–34.
- Mayer, J. & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D.J. Sluyter (Eds.), *Emotional development and emotional intelligence* (pp. 3–31). New York, NY: Basic Books.
- McClelland, D.C. (1973). Testing for competence rather than for intelligence. *American Psychologist*, *28*, 1–14. <http://doi.org/10.1037/h0034092>
- McClelland, D.C. (1985). *Human motivation*. Glenview, IL: Scott, Foresman.
- McClelland, D.C. & Boyatzis, R.E. (1982). Leadership motive pattern and long-term success in management. *Journal of Applied Psychology*, *67*, 737–743. <http://doi.org/10.1037/0021-9010.67.6.737>
- McClelland, D.C. & Franz, C.E. (1992). Motivational and other sources of work accomplishments in mid-life: A longitudinal study. *Journal of Personality*, *60*, 679–707. <http://doi.org/10.1111/j.1467-6494.1992.tb00270.x>
- McClelland, D.C., Koestner, R. & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, *96*, 690–702.
- McCormick, E.J., Jeanneret, P.R. & Mecham, R.C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology*, *56*, 347–368. <http://doi.org/10.1037/h0033099>
- McCrae, R.R. & Costa, P.T., Jr. (1997). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J.S. Wiggins (Ed.), *The Five-Factor Model of personality: Theoretical perspectives* (pp. 51–87). New York: Guilford Press.
- McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. & Maurer, S.D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, *79*, 599–616. <http://doi.org/10.1037/0021-9010.79.4.599>
- McFarland, L.A., Yun, G.J., Harold, C.M., Viera, L. & Moore, L.G. (2005). An examination of impression management use and effectiveness across assessment center exercises: The role of competency demands. *Personnel Psychology*, *58*, 949–980. <http://doi.org/10.1111/j.1744-6570.2005.00374.x>
- McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10. <http://doi.org/10.1016/j.intell.2008.08.004>

- Melchers, K. G., Ingold, P. V., Wilhelmy, A. & Kleinmann, M. (2015). Beyond validity: Shedding light on the social situation in employment interviews. In I. Nikolaou & J. K. Oostrom (Eds.), *Employee Recruitment, Selection, and Assessment: Contemporary Issues for Theory and Practice* (pp. 174–171). Hove, UK: Psychology Press.
- Meriac, J. P., Hoffman, B. J. & Woehr, D. J. (2014). A Conceptual and Empirical Review of the Structure of Assessment Center Dimensions. *Journal of Management*, 40, 1269–1296. <http://doi.org/10.1177/0149206314522299>
- Mischel, W. (1977). The Interaction of Person and Situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 333–352). Hillsdale: Lawrence Erlbaum.
- Moscoso, S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment*, 8, 237–247. <http://doi.org/10.1111/1468-2389.00153>
- Moser, K. (2007). Bewerbungsunterlagen und Referenzen. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 490–496). Göttingen: Hogrefe.
- Mount, M. K., Witt, L. A. & Barrick, M. R. (2000). Incremental validity of empirically keyed bio-data scales over GMA and the five factor personality constructs. *Personnel Psychology*, 53, 299–323. <http://doi.org/10.1111/j.1744-6570.2000.tb00203.x>
- Muchinsky, P. M. (1993). Validation of intelligence and mechanical aptitude tests in selecting employees for manufacturing jobs. *Journal of Business and Psychology*, 7, 373–382. <http://doi.org/10.1007/BF01013752>
- Murray, H. A. (1938). *Explorations in Personality*. New York: Oxford University Press.
- Ones, D. S. & Viswesvaran, C. (1996). Bandwidth – Fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17, 609–626. [http://doi.org/10.1002/\(SICI\)1099-1379\(199611\)17:6<609::AID-JOB1828>3.0.CO;2-K](http://doi.org/10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K)
- Ones, D. S. & Viswesvaran, C. (2001a). Personality at Work. Criterion-Focused Occupational Personality Scales Used in Personnel Selection. In B. W. Roberts & R. Hogan (Eds.), *Personality Psychology in the Workplace* (pp. 63–92). Washington, DC: American Psychological Association.
- Ones, D. S. & Viswesvaran, C. (2001b). Integrity tests and other criterion focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9, 31–39.
- Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome. Issues in organization and management series*. Lexington, MA: Lexington Books.
- Ott, M., Ulfert, A. & Kersting, M. (2017). „Online-Assessments“ und „Self-Assessments“ in der Eignungsdiagnostik. In D. E. Krause (Hrsg.), *Personalauswahl* (S. 215–242). Wiesbaden: Springer-Gabler.
- Otto, J. H., Döring-Seipel, E., Grebe, M. & Lantermann, E.-D. (2001). Entwicklung eines Fragebogens zur Erfassung der wahrgenommenen emotionalen Intelligenz. *Diagnostica*, 47, 178–187. <http://doi.org/10.1026//0012-1924.47.4.178>
- Pässler, K., Beinicke, A. & Hell, B. (2015). Interests and intelligence: A meta-analysis. *Intelligence*, 50, 30–51. <http://doi.org/10.1016/j.intell.2015.02.001>

- Paunonen, S. V. & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, *81*, 524–539. <http://doi.org/10.1037/0022-3514.81.3.524>
- Ployhart, R. E., Lim, B.-C. & Chan, K. Y. (2001). Exploring relations between typical and maximum performance ratings and the five factor model of personality. *Personnel Psychology*, *54*, 809–843. <http://doi.org/10.1111/j.1744-6570.2001.tb00233.x>
- Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M. & Blume, B. D. (2009). Individual- and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *The Journal of applied psychology*, *94*, 122–141. <http://doi.org/10.1037/a0013079>
- Podsakoff, P. M., MacKenzie, S. B., Paine, J. B. & Bacharach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, *26*, 513–563. <http://doi.org/10.1177/014920630002600307>
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322–338. <http://doi.org/10.1037/a0014996>
- Pulakos, E. D., Arad, S., Donovan, M. A. & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, *85*, 612–624. <http://doi.org/10.1037/0021-9010.85.4.612>
- Reddock, C. M., Biderman, M. D. & Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. *International Journal of Selection and Assessment*, *19*, 119–131. <http://doi.org/10.1111/j.1468-2389.2011.00540.x>
- Richardson, M., Abraham, C. & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*, 353–387. <http://doi.org/10.1037/a0026838>
- Robertson, I. T. & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, *74*, 441–472. <http://doi.org/10.1348/096317901167479>
- Roest, F., Scherzer, A., Urban, E., Gangl, H. & Brandstätter, C. (1989). *Mailbox '90. Ein computergestütztes Test- und Trainingsverfahren zur Personalentwicklung*. Wien und Weinheim: Scicon und Beltz.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F. & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. <http://doi.org/10.1016/j.intell.2015.09.002>
- Sackett, P. R., Zedeck, S. & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*, 482–486. <http://doi.org/10.1037/0021-9010.73.3.482>
- Salgado, J. F., Anderson, N. R., Moscoso, S., Bertua, C. & De Fruyt, F. (2003a). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*, 573–605.
- Salgado, J. F., Anderson, N. R., Moscoso, S., Bertua, C., De Fruyt, F. & Rolland, J. P. (2003b). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology*, *88*, 1068–1081.

- Salgado, J. F. & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology, 11*, 299–324. <http://doi.org/10.1080/13594320244000184>
- Schein, V.E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues, 57*, 675–688. <http://doi.org/10.1111/0022-4537.00235>
- Scherm, M. (Hrsg.). (2005). *360-Grad-Beurteilungen*. Göttingen: Hogrefe.
- 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. <http://doi.org/10.1037/0033-2909.124.2.262>
- Schmidt, F.L. & Hunter, J.E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162–173. <http://doi.org/10.1037/0022-3514.86.1.162>
- Schmidt, F.L., Law, K., Hunter, J.E., Rothstein, H.R., Pearlman, K. & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology, 78*, 3–12. <http://doi.org/10.1037/0021-9010.78.1.3>
- Schmidt, F.L. & Zimmerman, R.D. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology, 89*, 553–561. <http://doi.org/10.1037/0021-9010.89.3.553>
- Schmukle, S. C. & Egloff, B. (2011). Indirekte Verfahren zur Erfassung von Persönlichkeit („Objektive Persönlichkeitstests“). In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Persönlichkeitsdiagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Bd. 4, S. 73–120). Göttingen: Hogrefe.
- Schneider, B., Goldstein, H. W. & Smith, D. B. (1995). The ASA framework: An update. *Personnel Psychology, 48*, 747–773. <http://doi.org/10.1111/j.1744-6570.1995.tb01780.x>
- Schneider, W.J. & McGrew, K.S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test, and issues* (3rd ed., pp. 99–144). New York: Guilford Press.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie, 34*, 184–191.
- Schuler, H. (1992). Das Multimodale Einstellungsinterview. *Diagnostica, 38*, 281–300.
- Schuler, H. (2014). *Psychologische Personalauswahl. Eignungsdiagnostik für Personalentscheidungen und Berufsberatung* (4., vollst. überarb. und erw. Aufl.). Göttingen: Hogrefe.
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramit, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie, 6*, 60–70. <http://doi.org/10.1026/1617-6391.6.2.60>
- Scullen, S.E., Mount, M.K. & Judge, T.A. (2003). Evidence of the construct validity of developmental ratings of managerial performance. *Journal of Applied Psychology, 88*, 50–66. <http://doi.org/10.1037/0021-9010.88.1.50>
- Sende, C. & Moser, K. (2017). Zankapfel Arbeitszeugnis. Warum wir mehr Empirie in die juristische und mediale Debatte einbringen sollten. *Psychologische Rundschau, 68*, 131–134.
- Smith, P.C. & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149–155. <http://doi.org/10.1037/h0047060>

- Spangler, W.D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, *112*, 140–154. <http://doi.org/10.1037/0033-2909.112.1.140>
- Steinmayr, R., Schütz, A., Hertel, J. & Schröder-Abé, M. (2011). *Mayer-Salovey-Caruso Test zur Emotionalen Intelligenz (MSCEIT™)*. *Deutschsprachige Adaptation des Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT™) von John D. Mayer, Peter Salovey und David R. Caruso*. Bern: Huber.
- Sternberg, R.J. & Wagner, R.K. (Eds.). (1986). *Practical intelligence: Nature and origins of competence in the everyday world*. New York: Cambridge University Press.
- Stevens, C.K. & Kristof, A.L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology*, *80*, 587–606. <http://doi.org/10.1037/0021-9010.80.5.587>
- Strobel, A., Franke-Barthold, L. & Kersting, M. (2017). Eignungsinterview. In Diagnostik- und Testkuratorium (Hrsg.), *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 65–83). Heidelberg: Springer.
- Su, R., Rounds, J. & Armstrong, P.I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*, 859–884. <http://doi.org/10.1037/a0017364>
- Süß, H.-M. & Beauducel, A. (2011). Intelligenztests und ihre Bezüge zu Intelligenztheorien. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Bd. 3, S. 97–234). Göttingen: Hogrefe.
- Taylor, F.W. (1911). *The Principles of Scientific Management*. New York and London: Harper & Brothers.
- Tett, R.P. & Burnett, D.D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*, 500–517. <http://doi.org/10.1037/0021-9010.88.3.500>
- Tett, R.P. & Guterman, H.A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*, 397–423. <http://doi.org/10.1006/jrpe.2000.2292>
- Tett, R.P., Jackson, D.N. & Rothstein, M.G. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703–742. <http://doi.org/10.1111/j.1744-6570.1991.tb00696.x>
- Thorsteinson, T.J. & Balzer, W.K. (1999). Effects of coworker information on perceptions and ratings of performance. *Journal of Organizational Behavior*, *20*, 1157–1173. [http://doi.org/10.1002/\(SICI\)1099-1379\(199912\)20:7<1157::AID-JOB947>3.0.CO;2-J](http://doi.org/10.1002/(SICI)1099-1379(199912)20:7<1157::AID-JOB947>3.0.CO;2-J)
- Thunsdorff, C., Michaelis, L., Weis, S., Kersting, M., Schmitt, M. & Ulfert, A.-S. (2015). Studieneignung für spezifische Studienfächer – welche Rolle spielt die Auswahl der Anforderungsanalyseverfahren bei der Identifizierung erfolgsrelevanter Eigenschaften? *Das Hochschulwesen*, *63*, 98–103.
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, *21*, 11–27. <http://doi.org/10.1024/1010-0652.21.1.11>

- Truxillo, D. M., Donahue, L. M. & Kuang, D. (2004). Work samples, performance tests, and competency testing. In J. C. Thomas & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 4, pp. 345–370). Hoboken, NJ: Wiley.
- Van Iddekinge, C. H., Roth, P. L., Putka, D. J. & Lanivich, S. E. (2011). Are you interested? A Meta-Analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, *96*, 1167–1194. <http://doi.org/10.1037/a0024343>
- Van Iddekinge, C. H., Roth, P. L., 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. <http://doi.org/10.1037/a0021196>
- Vernon, P. (1959). *The structure of human abilities*. London: Methuen.
- Vinchur, A. J., Schippmann, J. S., Switzer, F. S. & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, *83*, 586–597. <http://doi.org/10.1037/0021-9010.83.4.586>
- Viswesvaran, C. & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, *8*, 216–226. <http://doi.org/10.1111/1468-2389.00151>
- Weis, S., Seidel, K. & Süß, H.-M. (2006). Messkonzepte sozialer Intelligenz – Literaturübersicht und Ausblick. In R. Schulze, P. A. Freund & R. D. Roberts (Hrsg.), *Emotionale Intelligenz: Ein Internationales Handbuch* (S. 213–234). Göttingen: Hogrefe.
- Wiesner, W. & Cronshaw, S. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, *61*, 275–290. <http://doi.org/10.1111/j.2044-8325.1988.tb00467.x>
- Williams, J. E. & Best, D. L. (1990). *Sex and the Psyche: Gender, Roles and Self Concepts viewed cross culturally*. Beverly Hills, CA: Sage.
- 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*, 601–617. <http://doi.org/10.1177/014920639101700305>
- Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., S. 505–560). New York: Plenum.
- Yukl, G. A. (2012). *Leadership in organizations* (8th ed.). London: Prentice-Hall.
- Zettler, I. & Hilbig, B. E. (2010). Honesty-Humility and a person-situation interaction at work. *European Journal of Personality*, *24*, 569–582. <http://doi.org/10.1002/per.757>
- Zimmer, B. & Sander, M. (2014). Berufswahl: Können Interessen Fähigkeiten ersetzen? *Wirtschaftspsychologie aktuell*, *2*, 41–46.

ENZYKLOPÄDIE DER PSYCHOLOGIE

In Verbindung mit der
Deutschen Gesellschaft für Psychologie

herausgegeben von

Prof. Dr. Niels Birbaumer, Tübingen
Prof. Dr. Dieter Frey, München
Prof. Dr. Julius Kuhl, Osnabrück
Prof. Dr. Wolfgang Schneider, Würzburg
Prof. Dr. Ralf Schwarzer, Berlin

Themenbereich B

Methodologie und Methoden

Serie III

Psychologische Interventionsmethoden

Band 3

Methoden der

Arbeits-, Organisations- und
Wirtschaftspsychologie

 hogrefe

Methoden der Arbeits-, Organisations- und Wirtschaftspsychologie

2018

herausgegeben von

Prof. Dr. Siegfried Greif, Osnabrück
apl. Prof. Dr. Kai-Christoph Hamborg, Osnabrück

 hogrefe