

Managementdiagnostik: Verfahren und Qualitätsaspekte

Martin Kersting

Stichwörter

Potentialbeurteilung, Personalauswahl, Eignungsdiagnostik, Managementdiagnostik, Bewerbungsunterlagen, Lebenslaufanalyse, Bewerberportal, Interview, Interviewleitfaden, Rollenspiel, Gruppendiskussion, Präsentationsübung, Case-Study, Postkorb, Test, Intelligenztest, Persönlichkeitsfragebogen, Assessment-Center, Development-Center, Audit, DIN 33430, Qualität, Qualitätssicherung

Abstract

Der Artikel stellt häufig genutzte Verfahren der Potentialfeststellung und (internen sowie externen) Personalauswahl wie Interviews, Rollenspiele, Persönlichkeitsfragebogen, Leistungstests und Assessment Center kurz vor und klassifiziert diese in ein übergeordnetes System (CUBE). Das CUBE System erleichtert es, das Erfolgsprinzip des Verfahrenspluralismus zu verstehen und umzusetzen. Für die vorgestellten Verfahren werden Empfehlungen für die Praxisanwendung sowie Qualitätskriterien formuliert. Dem Überblick ist ein kurzer Abschnitt über die Anforderungsorientierung als Voraussetzung jeglichen Verfahrenseinsatzes vorangestellt.

1. Potentialfeststellung und (interne sowie externe) Personalauswahl

“Innovation has nothing to do with how many R&D dollars you have. When Apple came up with the Mac, IBM was spending at least 100 times more on R&D. It’s not about money. It’s about the people you have (...).”

Steve Jobs, Fortune, Nov. 9, 1998

Gutes Personal ist eine entscheidende Voraussetzung für den Erfolg einer Organisation. Faktoren wie die demographische Entwicklung (weniger Menschen) sowie eine misslungene Bildungsplanung (zu wenig qualifizierte Menschen) verschärfen den Wettbewerb um die Potentialträger. Auf diese zentrale Herausforderung sollte man mit einem strategisch ausgerichteten Gesamtkonzept reagieren, welches sowohl die Organisationsentwicklung (z.B. Imagebildung, Arbeitszeitmodelle, Unternehmens- und Führungskultur) als auch die Personalgewinnung, -auswahl und -entwicklung umfasst. (Zur Personalentwicklung siehe auch den Beitrag von Teuber unter Ziffer 3970).

Der vorliegende Beitrag fokussiert die Verfahren der Potentialfeststellung und (internen sowie externen) Personalauswahl, die hier unter dem Begriff der Managementdiagnostik zusammengefasst werden. Der Begriff soll dabei nicht, wie von Sarges (2006) definiert, allein auf den Personenkreis der Manager¹ angewendet werden, sondern in einem weiten Begriffsverständnis alle Personengruppen und somit die gesamte berufliche Eigenschaftsdiagnostik umfassen.

2. Anforderungsorientierung

Die Verfahren der Managementdiagnostik sind Hilfsmittel, um zielgerichtet und systematisch relevante Informationen über interne und externe Kandidaten zu sammeln, aufzubereiten und zu interpretieren. Bevor ein Verfahren ausgewählt oder konstruiert werden kann, muss geklärt werden, welche Informationen überhaupt relevant/erfolgskritisch sind. D.h. bevor man sich mit den Kandidaten auseinandersetzt und Informationen zum „Ist-Zustand“ erhebt, muss man sich im Rahmen einer Arbeits- und Anforderungsanalyse mit der in Frage stehenden Tätigkeit/Aufgabe auseinandersetzen. Erst nach dieser Auseinandersetzung können Anforderungen (Soll-Werte) an die Personen formuliert werden, die diese Tätigkeit erfolgreich ausüben bzw. der Aufgabe gerecht werden sollen. Bei den Anforderungen geht es sowohl um hard facts (z.B. definierte Abschlüsse, Erfahrungen), als auch um soft-skills. Die Anforderungsanalyse erfordert eine enge Abstimmung mit dem Auftraggeber. Wenn die letzte Phase der Personalberatung, die Präsentation der Kandidaten beim Auftraggeber, scheitert, liegt es häufig daran, dass in der ersten Phase, der Anforderungsanalyse, nicht gut gearbeitet wurde. Es reicht nicht aus, sich auf abstrakte Phrasen und Leerformeln zu beschränken (z.B. „wir suchen flexible high potentials mit Persönlichkeit und sozialer Kompetenz“). Vielmehr muss genau definiert werden, was unter den verwendeten Begriffen verstanden wird und welche konkret beobachtbaren Verhaltensweisen damit verbunden werden. Zur Anforderungsanalyse stehen verschiedene Verfahren zur Verfügung, etwa die „critical incident technique“ und die „repertory-grid-technique“. Einen Überblick geben z.B. Brannick und Levine (2002) sowie Schuler (2006).

Als Ergebnis der Anforderungsanalyse erhält man in der Regel ein Anforderungs- oder Kompetenzprofil, welches generische und/oder organisationsspezifische Anforderungsdimensionen/Kompetenzen umfasst. Dieses Profil ist dann praktisch nützlich, wenn die Anzahl der Anforderungsdimensionen/Kompetenzen nicht zu hoch ist, und wenn die Dimensio-

¹ Um schwer verständliche Satzkonstruktionen zu vermeiden, wird im Folgenden das generische Maskulinum verwendet, gleichwohl männliche und weibliche Personen in gleicher Weise gemeint sind.

nen/Kompetenzen untereinander trennscharf sind. Die Dimensionen/Kompetenzen müssen erfolgsrelevant sein und es muss eine verhaltensnahe Beschreibung vorliegen (z.B. mit negativen und positiven Verhaltensankern). Dabei sollte zwischen „must have“ Kriterien und „nice to have“ Kriterien unterschieden werden. Bei jeder Dimension sollte auch überlegt werden, ob eine Nichterfüllung der Anforderung bedeutet, dass die Person insgesamt ungeeignet ist, oder ob (und wie) Schwächen kompensiert werden können. Gute Personalbeurteilung ist stets systematisch mit der Personalentwicklung verknüpft. Umgekehrt formuliert: Berater, die sich darauf beschränken, Personen in ein neues Arbeitsverhältnis zu bringen und der Person sowie der Organisation keine konkreten Hinweise zur Weiterentwicklung geben, bleiben hinter den Möglichkeiten moderner Personalbeurteilung zurück.

Außerdem wird im Rahmen der Anforderungsanalyse auch der Bewertungsmaßstab erarbeitet. Für die Personalauswahl reicht es nicht aus, die Einzeldimensionen zu betrachten, sondern es muss vorab geregelt werden, wie das Insgesamt der Informationen zu einem Gesamturteil zusammengefügt wird (siehe auch den letzten Abschnitt des vorliegenden Beitrags). Aus einer gut durchgeführten Anforderungsanalyse ergeben sich in der Regel – quasi als Nebennutzen – zahlreiche Anregungen für die Gestaltung des Verfahrens, etwa Ideen für Interviewfragen und Rollenspiele.

Angesichts des Personalmangels muss man sich genau überlegen, wie hoch man die Anforderungs-Hürden setzen will. Weder darf man eine Mindestqualität unterschreiten, noch darf man künstliche Hürden durch die Formulierung von Anforderungen schaffen, die zwar gut klingen, de facto aber für den Erfolg irrelevant sind.

3. Verfahren/Datenquellen

Die Beurteilung von Menschen basiert letztendlich immer auf Daten. Die unterschiedlichen Verfahren können danach sortiert werden, welche Daten auf welche Art und Weise erhoben werden. Kersting (2006, 2010) hat die vorhandenen Verfahren im CUBE System nach deren Datenquelle sortiert. Demzufolge kann man

- etwas aus dem Leben der Person aufgrund objektiver (im Sinne von: verifizierbarer, also zumindest prinzipiell nachprüfbarer) Daten (Sachinformationen) in Erfahrung bringen, z.B. ihren Werdegang, ihre Leistungen analysieren
- die Zielperson selbst oder Dritte (z.B. Vorgesetzten, Kollegen, Kunden usw.) zur Zielperson befragen

- die Person in Situationen bringen/Übungen absolvieren lassen, die eine bestimmte Realität simulieren (z.B. im Rollenspiel)
- die Person (Selbsteinschätzung) oder eine Auskunftsperson (Fremdeinschätzung) mit Hilfe eines Fragebogens befragen,
- die Person testen.

Diese Differenzierung führt zu der folgenden Verfahrenseinteilung (siehe Abbildung 1):

1. Dokumenten- und Faktenanalysen (z.B. die Analyse und Interpretation von ((Hoch-)Schul-, Arbeits-) Zeugnissen, dem Lebenslauf, von Beurteilungen (auch 360-Grad Beurteilungen), von Internet-Funden über den Kandidaten usw.)
2. Direkte Befragungen (z.B. Interview, Gespräch mit einem Referenzgeber)
3. Verfahren zur Verhaltensbeobachtung (z.B. Rollenspiele, Gruppendiskussionen, Präsentationsübungen usw.)
4. Psychometrische Fragebogen (z.B. ein Persönlichkeitsfragebogen)
5. Psychometrische Leistungstests (z.B. ein Intelligenztest, ein Wissenstest, eine schriftliche Fachprüfung)

Bei den Kategorien (1), (2) und (3) erfolgt die Bewertung letztendlich durch einen Beurteiler, auch wenn sich diese Beurteilung natürlich an klaren Kriterien orientieren sollte und die Beurteiler ein Training durch-

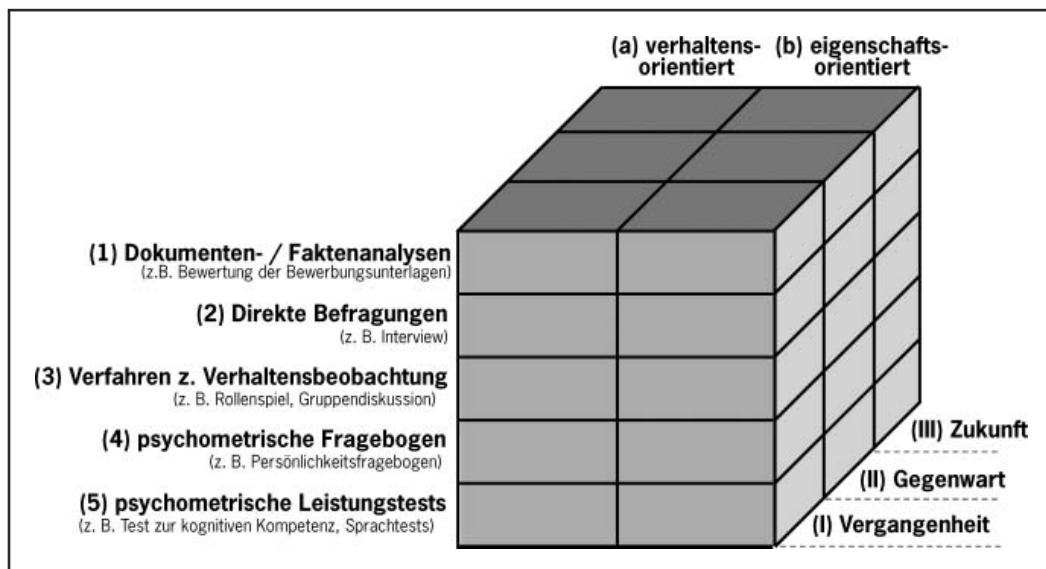


Abbildung 1: Das CUBE-System zur Klassifizierung unterschiedlicher Verfahren der Potentialeinschätzung und (internen sowie externen) Personalauswahl

laufen haben sollten. Bei den psychometrischen Verfahren (4) und (5) erfolgt die Bewertung hingegen meistens nach einem vorgegeben „Automatismus“, nach strengen Auswertungsregeln, die kaum Interpretationsspielräume lassen. Beispielsweise werden psychometrische Verfahren häufig über „Schablonen“ ausgewertet. Im Extremfall erfolgt die Auswertung durch einen Computer. Somit ist die Auswertungsobjektivität maximal hoch. Dieser Aspekt der Standardisierung entscheidet mit darüber, ob ein Verfahren in die Kategorien (1), (2) und (3) oder in die Kategorien (4) und (5) einzuordnen ist.

Eine Interpretation des Lebenslaufes durch einen Berater fällt in Kategorie (1), ein mittels Schablone ausgewerteter biographischer Fragebogen oder ein mittels Computer ausgewertetes Online-Bewerbungsportal (bei dem der Kandidat Angaben zu seinem Lebenslauf getroffen hat) fallen hingegen in Kategorie (4).

In die Kategorie (3), Verfahren zur Verhaltensbeobachtung, fallen Arbeitsproben, wenn die Bewertung aufgrund von Beobachtungen durch eine oder mehrere Person(en) vorgenommen wird. Es gibt aber auch Arbeitsproben, die vollkommen standardisiert (z.B. durch einen Computer) bewertet werden, diese Arbeitsproben gelten in der vorliegenden Einteilung als psychometrischer Leistungstest. Entsprechend verhält es sich mit den häufig eingesetzten Verfahren case-studies und Postkörben, die entweder lediglich Grundlage einer Präsentation sind (und entsprechend in Kategorie (3) eingeordnet werden) oder aber wie ein Leistungstest ausgewertet werden (Kategorie 5).

Der Unterschied zwischen der vierten (psychometrische Fragebogen) und der fünften Kategorie (Leistungstests) besteht darin, dass es im letzteren Fall eine „natürliche“, vom Kandidaten eindeutig erkennbare Bewertung gibt, z.B. bei Tests zur kognitiven Kompetenz oder bei Wissenstests nach „richtig“ oder „falsch“, bei Tests zur Bearbeitungsgeschwindigkeit nach „schnell“ oder „langsam“ (siehe unten).

Assessment Center/Development Center, Management-Audits usw. kommen in der Aufstellung der Verfahren nicht vor, da mit diesen Begriffen kein einzelnes Verfahren, sondern ein Methoden-Mix bezeichnet wird (siehe die Ausführungen zum Verfahrenspluralismus weiter unten). Das CUBE System leistet eine Differenzierung der Verfahren. Ein Management-Audit umfasst z.B. häufig ein Interview (Kategorie 2) sowie ein Rollenspiel oder eine Präsentationsübung (Kategorie 3). Ein Assessment oder Development Center umfasst überwiegend Simulationsaufgaben (wie z.B. Rollenspiele, Gruppendiskussionen, Kategorie 3), aber in der Regel auch ein Interview (Kategorie 2) und – im Falle von qualitativ hochwertigen Assessment/Development Centern – auch psychometrische Verfahren (Kategorien 4 und 5).

Das CUBE System von Kersting (2006, 2010, siehe Abbildung 1) sieht noch zwei weitere Dimensionen zur Klassifikation von Verfahren vor. Mit der zweiten Dimension des CUBE Systems wird unterschieden, ob das Verfahren primär auf die Erfassung von Verhalten oder auf Eigenschaften zielt.

Die Zeitdimension stellt die dritte und letzte Dimension des CUBE Systems dar. Hier werden die Verfahren danach eingeteilt, ob sie sich primär an der Vergangenheit, der Gegenwart oder der Zukunft orientieren. Auf diese beiden Dimensionen wird im vorliegenden Beitrag nur am Rande eingegangen, der Artikel konzentriert sich auf die fünf Ausprägungen der ersten CUBE Dimension (Verfahren) und zählt diese entsprechend als „1–1“ bis „1–5“ (Dimension 1, Ausprägung 1 (Dokumenten- und Faktenanalysen) bis Dimension 1, Ausprägung 5 (psychometrische Leistungstests)).

Im Folgenden werden die einzelnen Verfahrensgruppen kurz anhand typischer Beispiele erläutert und es werden Praxis-Empfehlungen zur Verfahrensanwendung formuliert. Einleitend werden Qualitätskriterien genannt, die verfahrensübergreifend gelten.

4. Allgemeine Qualitätskriterien für Verfahren

Jedes Verfahren der Managementdiagnostik muss (1) objektiv, (2) zuverlässig und (3) gültig sein.

Objektivität meint, dass das Verfahrensergebnis unabhängig von den Personen ist, die das Verfahren (a) durchführen, (b) auswerten und das Verfahrensergebnis (c) interpretieren.

Um die Frage des Messfehlers geht es bei dem Qualitätsgesichtspunkt der Zuverlässigkeit (Reliabilität). Mit den Verfahren werden Messungen vorgenommen, jede Messung ist mit einem Fehler behaftet. Umso kleiner der Messfehler, umso größer die Zuverlässigkeit. Um die Zuverlässigkeit eines Verfahrens zu bestimmen, wird der Zusammenhang zwischen zwei oder mehreren Werten, die dasselbe erfassen, empirisch berechnet. Zum Beispiel: Wie hoch stimmen die Erst- und die Zweitbeurteilung zu stabilen Persönlichkeitsmerkmalen überein, wenn das Interview mit dem Kandidaten nach vier Wochen wiederholt wird?

Der dritte und bedeutsamste Qualitätsaspekt betrifft die Gültigkeit (Validität). Dabei geht es streng genommen nicht um das Verfahren, sondern um die Aussagekraft der aus dem Verfahren abgeleiteten Schlussfolgerungen. Ist ein Kandidat, dem aufgrund eines Interviews eine hohe Sozialkompetenz zugeschrieben wird, wirklich sozial kompetent oder nur sprachlich gewandt? (Konstruktvalidität). Lässt sich anhand des Interviews mit dem Kandidaten wirklich der spätere Erfolg im Vertrieb vorher-

sagen (Kriteriumsvalidität)? Dies ist anhand von Bewährungskontrollen empirisch zu prüfen.

Die drei Qualitätskriterien stehen in einem Abhängigkeitsverhältnis zueinander (siehe Abbildung 2). Ein Verfahren, das keine Objektivität aufweist, kann nicht zuverlässig und gültig sein. Die Zuverlässigkeit wiederum ist eine notwendige Voraussetzung für die Gültigkeit. Deshalb ist es wichtig, so viel Wert auf die Objektivität zu legen. Ob ein Verfahren objektiv ist, lässt sich zumindest in einer ersten Annäherung leicht beurteilen: Hängt das Ergebnis des Verfahrens (z.B. die Bewertung der Bewerbungsunterlagen, der Interviewleistungen, der Ergebnisse im Persönlichkeitsfragebogen usw.) wirklich nur von den Fakten oder der Leistung des Kandidaten ab oder spielt es auch eine Rolle, welcher Berater sich die Bewerbungsunterlagen anschaut, das Interview führt oder beurteilt sowie den Persönlichkeitsfragebogen auswertet und interpretiert? Im letzteren Falle sinkt die Objektivität und in Folge die Zuverlässigkeit und Gültigkeit.

Über diese drei hauptsächlichen Qualitätskriterien hinaus gibt es weitere (Neben-)Kriterien, etwa die Akzeptanz (auch soziale Validität genannt, siehe z.B. Kersting, 2008b).

Verfahrensübergreifend gilt, dass eine qualitativ hochwertige Managementdiagnostik in jedem Fall ein Feedback über die Verfahrensergebnisse an den Kandidaten vorsieht.

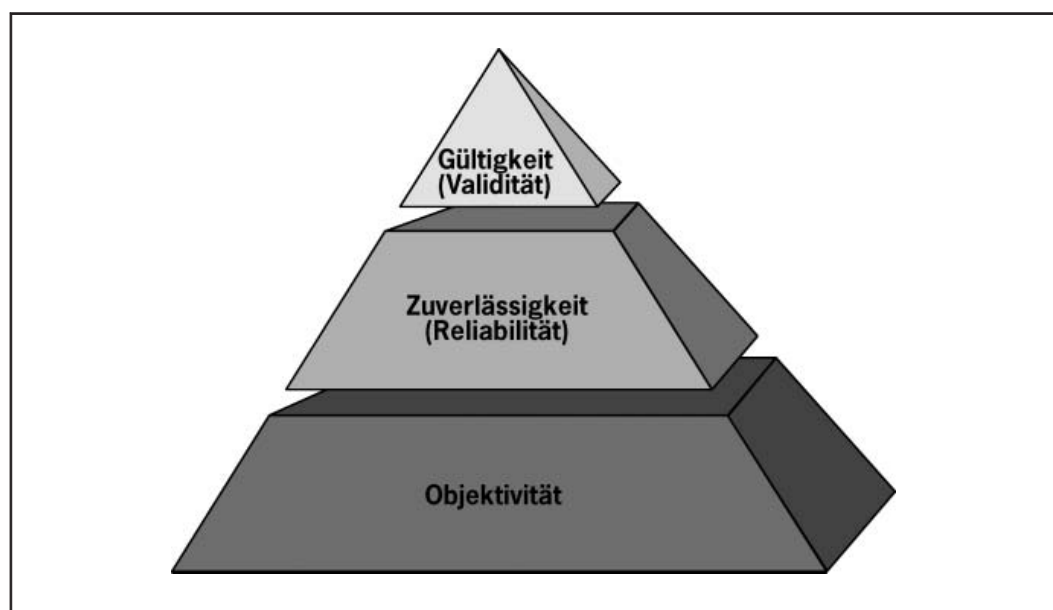


Abbildung 2: Allgemeine Qualitätskriterien für Verfahren der Potentialeinschätzung und (internen sowie externen) Personalauswahl

5. Verfahren der Potentialfeststellung und (internen sowie externen) Personalauswahl

5.1 CUBE Dimension 1–1: Dokumenten- und Faktenanalysen (z. B. Sichtung von Bewerbungsunterlagen)

Bei den Verfahren der Dokumenten- und Faktenanalyse geht es darum, „Spuren“ menschlichen Verhaltens (von Lebenslaufdaten über Zeugnisse bis hin zu einzelnen Vorzeige-Projekten) zu sichten und zu interpretieren. Ein wichtiges Merkmal dieser Daten ist deren Historizität. Die Praxis setzt vor allem auf die Auswertung von Bewerbungsunterlagen und Lebensläufen. Diese gelingt umso besser, wenn sie durch ein klares Regelwerk geleitet und im Idealfall von zwei unabhängig voneinander arbeitenden Personen vorgenommen wird. Häufig wird die Dokumenten- und Faktenanalyse an Assistenzkräfte delegiert. Dagegen ist nichts einzuwenden, sofern derart eindeutige Auswertungsvorgaben bestehen, dass die Assistenzkraft zur gleichen Entscheidung kommt wie die verantwortliche Person. Zu regeln ist z. B., wie diskontinuierliche Lebensläufe zu bewerten sind. Sofern die Auswertungsobjektivität nicht gewährleistet wird – und das ist häufig der Fall – hat die Assistenzkraft letztendlich eine Weichenfunktion. Der Verantwortliche sieht und begutachtet nur noch die Personen, deren Unterlagen die Assistenzkraft überzeugt haben. Ein solches Vorgehen kann den Fehler der zweiten Art begünstigen, der darin besteht, einen potentiell geeigneten Kandidaten zu übersehen. In Zeiten des Personalmangels kommt es verstärkt darauf an, möglichst viele gute Bewerber zu „entdecken“ (gesteigerte Sensitivität).

Eine interessantes Hilfsmittel im Rahmen der Dokumenten- und Faktenanalysen sind Bewerberportale, bei denen die Kandidaten selbst Einträge zu Ihren Zeugnissen, zum Lebenslauf usw. vornehmen. Um eine computergestützte Auswertung zu ermöglichen, muss die Bewertung sowohl der einzelnen Aspekte als auch der Kombination von Informationen vorab geregelt werden. Dies zwingt die Verantwortlichen, ihre diesbezüglichen Hausaufgaben zu erledigen. Außerdem ersparen solche Systeme viel Zeit, verkürzen die Response-Zeiten und erleichtern die Dokumentation und Speicherung der Daten. Die notwendige Überprüfung der Selbstauskünfte kann zu einem späteren Zeitpunkt erfolgen und sich auf diejenigen Personen beschränken, die in die engere Wahl kommen.

5.2 CUBE Dimension 1–2: Direkte Befragungen (z. B. Interview)

Aus der Gruppe der direkten Befragungen soll hier das Interview mit dem Kandidaten als häufig angewendetes Verfahren herausgegriffen werden. Qualitativ hochwertige Interviews zeichnen sich durch ein hohes Maß an Strukturierung aus. In diesem Sinne vorbereitete und kontrolliert durchgeführte sowie ausgewertete Interviews sind anderen Interviews

an Treffsicherheit deutlich überlegen (McDaniel, Whetzel, Schmidt und Maurer, 1994). Grundlage eines derartigen Interviews ist ein Interviewleitfaden. Zumindest eine Teilmenge der im Interview zu stellenden Fragen und ihre Reihenfolge sollten vorbereitet sein. Dabei sollten – sofern anforderungsanalytisch geboten – unterschiedliche Fragearten, beispielsweise situative und biographische Fragen zum Einsatz kommen. Mit biographieorientierten Fragen wird die Vergangenheit und Gegenwart thematisiert und mit situativen Fragen („Stellen Sie sich bitte die folgende Situation vor ...“) die Zukunft. Das Beispiel illustriert auch die Zeitkategorie als dritte Dimension der CUBE Klassifikation (siehe Abbildung 1). Sowohl biographische als auch situative Fragen gehören zur Gruppe der direkten Befragung (CUBE Dimension 1–2), die beiden Fragetypen unterscheiden sich aber auf der dritten CUBE Dimension, der Zeitdimension (nämlich CUBE Dimension 3-I sowie 3-II für biographische Fragen und CUBE Dimension 3-III für situative Fragen).

Im Rahmen der Interviewvorbereitung sollte man sich nicht nur Gedanken über die Fragen, sondern auch über die Bewertung möglicher Antworten machen: Welche Antwort stellt eine gute oder schlechte Reaktion auf die Frage dar, welche Antwort wird eigentlich erwartet?

Eine Strukturierung erhöht maßgeblich die Vergleichbarkeit der Interviews sowohl über verschiedene Interviewte als auch über verschiedene Interviewer hinweg. Gut vorbereitete und kontrolliert abgewickelte Interviews weisen einen höheren Anforderungsbezug und eine geringere Manipulierbarkeit seitens der Kandidaten auf. Die Vorbereitung trägt außerdem zur Entlastung der Interviewer bei. Die Gefahr, dass der Interviewer mehr spricht als der Interviewte, sinkt.

So sinnvoll die Interviewvorbereitung und die Erarbeitung eines Gesprächsleitfadens ist, so wenig hilfreich ist es allerdings, die einzelnen Punkte einfach „abzuhaken“, ohne wirklich zuzuhören und ohne zu prüfen, ob die bislang thematisierten Aspekte hinreichend ausgeführt wurden. Daher sollte man häufig Nachfragen stellen. Eine Hilfe zur Formulierung von Nachfragen liefert das Verhaltens- oder STAR Dreieck (STAR für *SituaTion*, *Aktion* und *Resultat*): Demnach bedarf jede vollständige Schilderung (1) einer Situation/Aufgabe oder Ausgangslage (2), einer Handlung und (3) eines Resultats. Bei jeder Aussage des Kandidaten wird gedanklich geprüft, ob alle drei Seiten des Dreiecks angesprochen wurden, ansonsten wird eine entsprechende Nachfrage gestellt, z.B. „Welches Ergebnis haben Sie damit erzielt?“ für den Fall, dass das Resultat unklar blieb usw. Eine Variante dieser Technik ist die SOARA Interview-Technik (Situation, Objective, Action, Results, Aftermath).

Ein weiterer Erfolgsfaktor für das Interview ist die Protokollierung und Dokumentation. Das Protokollieren und Dokumentieren zählt zu den In-

interviewerpflichten, die in der Praxis häufig sträflich vernachlässigt werden.

Interviewer sollten sich regelmäßig Trainings und/oder Supervisionen unterziehen.

Einen Überblick über Theorie und Praxis von Interviews geben u.a. Schuller (2002) sowie Strobel und Westhoff (2011). Der Arbeitskreis Assessment Center hat Interview-Standards herausgegeben (siehe <http://www.arbeitskreis-ac.de>).

5.3 CUBE Dimension 1–3: Verfahren zur Verhaltensbeobachtung (z. B. Rollenspiele, Gruppendiskussionen, Präsentationen)

Vielen Verfahren der Personalbeurteilung wohnt eine „indirekte“ Logik inne. Hat jemand in der Vergangenheit hervorragendes geleistet (laut Dokumenten- und Faktenanalyse), so wird er – so die Schlussfolgerung – auch in Zukunft ein Leistungsträger sein. Beweist jemand in einer komplizierten abstrakten Intelligenzaufgabe (psychometrischer Test) sein Denkvermögen, so wird er – so die Schlussfolgerung – auch den kognitiven Herausforderungen des Berufs gerecht werden. Bei diesen Verfahren ist das beobachtete (Ankreuz-)Verhalten lediglich ein Mittel zum Zweck. Es dient dazu, auf ein „dahinter liegendes“, nicht direkt beobachtbares Konstrukt (die Anforderungsdimension/Kompetenz), z.B. auf die Leistungsmotivation oder die Intelligenz des Kandidaten zu schließen. Die Verfahren liefern dieser Auffassung zufolge „signs“, Indikatoren für das eigentlich interessierende Konstrukt. Bei den Verfahren zur Verhaltensbeobachtung ist hingegen das beobachtbare Verhalten bereits das Ziel. Durch die Gestaltung von Situationen wird ein Verhalten provoziert, welches unmittelbar ein Beispiel oder eine Stichprobe (sample) für die letztendlich interessierenden Verhaltensweisen ist, der „Umweg“ über ein Konstrukt ist dieser Auffassung zufolge nicht notwendig (siehe die Darstellung des „sign“ und „sample“ Ansatzes bei Wernimont und Campbell, 1968). Auch situative Fragen im Interview („Stellen Sie sich bitte die folgende Situation vor ...“) korrespondieren mit der Grundidee des sample-Ansatzes.

Ein Kandidat für den Vertrieb wird gebeten, eine kleine Produktpräsentation zu leisten; ein Kandidat für eine Führungsposition soll in einem Rollenspiel ein Gespräch mit einem schwierigen Mitarbeiter führen usw. Die Beobachter können dann – so die Annahme – unmittelbar das interessierende, erfolgsrelevante Verhalten beobachten. Das Verhalten in diesen Situationen oder „Übungen“ (z.B. Präsentationsaufgabe, Rollenspiel, Gruppendiskussion) wird dann regelgeleitet beobachtet und beurteilt, häufig von mehreren, vorab geschulten Personen. Einen Überblick zum Thema Verhaltensbeobachtung geben u.a. Spinath und Becker (2011). Qualitäts-

standards für Verfahren zur Verhaltensbeobachtung werden u. a. vom Arbeitskreis Assessment Center (2004, siehe <http://www.arbeitskreis-ac.de>) und in der DIN Norm zur Personalauswahl (DIN 33430, DIN, 2002, siehe Kersting, 2008a) formuliert. Kersting (2008c) hebt u. a. die folgenden Erfolgsfaktoren hervor:

- Die Übungskonstruktion sollte anforderungsorientiert, erfolgskritisch und realitätsorientiert (Simulationsprinzip) erfolgen.
- Die Übungskonstruktion sollte dimensionsorientiert erfolgen. (Ist die Übung geeignet, das gewünschte Verhalten, zum Beispiel „Durchsetzungsstärke“, zu provozieren?)
- Pro Übung (z.B. Rollenspiel) sollten maximal drei bis fünf Anforderungsdimensionen/Kompetenzen bewertet werden.
- Die Materialien und Instruktionen müssen sorgfältig gestaltet werden.
- Bei Rollenspielen muss es ausgearbeitete „Drehbücher“ und Trainings für die Rollenspieler geben.
- Die Übungen dürfen weder zu leicht, noch zu schwer sein, müssen eine angemessene Schwierigkeit haben.

Hinsichtlich der Verhaltensbeobachtung und -beurteilung fordert die DIN-Norm zur Personalauswahl (DIN 33430, DIN, 2002, siehe Kersting, 2008a) u. a., dass die Beobachtung und Beurteilung vorab geregelt werden muss. Wenn mehrere Beobachter ein Urteil abgeben, muss der Grad der Übereinstimmung zwischen den beurteilenden Personen berechnet werden (DIN-Screen Aussage 210 nach Kersting 2008a). An diesem Beispiel lässt sich besonders gut demonstrieren, wie wichtig Feedback- und Lernprozesse für die Managementdiagnostik sind. Bestimmt man z.B. unmittelbar nach einem Rollenspiel die Übereinstimmung der Beurteiler, so stellt man vielleicht fest, dass die Beurteiler sich in drei Beurteilungsdimensionen in der Regel sehr einig sind, in einer vierten aber nicht. Die Diskrepanz zwischen Ist- und Soll-Wert wird unmittelbar messbar. Durch dieses Feedback wird nun ein Optimierungsprozess ausgelöst. Die Beurteiler tauschen sich erneut über ihr Verständnis der kritischen Dimension und des Beurteilungsmaßstabes aus, legen gemeinsam neue Operationalisierungen fest usw. Ohne das Feedback über die Beurteilerübereinstimmung wäre das Defizit des Vorgehens hingegen nicht aufgefallen.

Verfahren zur Verhaltensbeobachtung können technisch unterstützt werden, etwa durch den Einsatz von Aufnahmegegeräten und vor allem durch Hilfsmittel zur Strukturierung des Abstimmungsprozesses zwischen mehreren Beurteilern.

5.4 CUBE Dimension 1–4: Psychometrische Fragebogen (z. B. ein Persönlichkeitsfragebogen)

U.a. aufgrund der Relevanz von Persönlichkeitsmerkmalen für die Prognose von Trainings- und Berufserfolg sowie Berufszufriedenheit werden auch Persönlichkeitsfragebogen zur Personalauswahl und -entwicklung eingesetzt (für einen Überblick siehe z.B. Hossiep & Mühlhaus, 2005). Die meisten Fragebogen nutzen Selbsteinschätzungen, wobei unterschiedliche Antwortformate vorgesehen werden. Während früher unspezifische Fragebogen zum Einsatz kamen, stehen in jüngerer Zeit auch berufsbezogene Inventare zur Verfügung. Berufsbezogene Persönlichkeitsinventare unterscheiden sich von den unspezifischen Persönlichkeitsinventaren u.a. dadurch, dass die „Reize“ (z.B. die Fragen) einen berufsbezogenen Kontext vorsehen. Die unspezifischen oder allgemeinen Persönlichkeitsinventare nutzen als Reiz (Items) entweder ausformulierte Sätze (z.B. „Ich esse gerne.“) oder aber Adjektive (z.B. „Sind Sie eher ‚dominant‘ oder ‚gewissenhaft?‘“). Das Beispiel illustriert auch die zweite Dimension der CUBE Klassifikation (siehe Abbildung 1), die Unterscheidung zwischen einer verhaltensorientierten und einer eigenschaftsorientierten Ausgestaltung eines Verfahrens. Die beiden Beispielitems „Ich esse gerne“ und „Sind Sie eher ‚dominant‘ oder ‚gewissenhaft?‘“ gehören beide zur Gruppe der psychometrischen Fragebogen, (CUBE Dimension 1–4), die beiden Fragetypen unterscheiden sich aber auf der zweiten CUBE Dimension (nämlich CUBE Dimension 4-a für die verhaltensnahe Formulierung und CUBE Dimension 4-b für das Adjektiv-Item). Ein berufsspezifischer Kontext kann natürlich eher bei Persönlichkeitsinventaren aktiviert werden, bei denen die Items nicht nur aus einzelnen Adjektiven, sondern aus ganzen Sätzen bestehen (z.B. „In beruflichen Konfliktsituationen ist es mir wichtig, einen Kompromiss zu finden.“). Mit berufsbezogenen Formulierungen arbeiten u.a. das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP, Hossiep und Paschen, 2003), das Leistungsmotivationsinventar (LMI, Schuler und Prochaska, 2000) und der Occupational Personality Questionnaire (SHL, 2000). Darüber hinaus sehen berufsbezogene Persönlichkeitsinventare Vergleichsdaten (Normen) von fragestellungsrelevanten Gruppen (z.B. Bewerbern) vor.

So genannte „Typen-Tests“ als Persönlichkeitsfragebogen sind zwar beliebt, orientieren sich häufig aber an psychoanalytischen Theorien, die gegenwärtig im wissenschaftlichen Bereich kaum noch genutzt und beforscht werden. Einige Typentests gehen beispielsweise auf antiquierte Theorien von Jung oder Marston zurück, für deren Gültigkeit bis heute keine empirischen Belege erbracht werden konnten (Jäger, 2004). Während mit Typentests, entsprechend dem vor-empirischen Stand der damaligen Wissenschaft, häufig lediglich die grobe (bipolare) Zugehörigkeit einer Person zu einem nur aus Plausibilitätsgründen konstruierten „Typ“

festgelegt wird, ist es mit dimensionsorientierten Persönlichkeitsfragebogen zu aktuellen Theorien möglich, die Ausprägung einer Persönlichkeitseigenschaft auf verschiedenen, mehr oder minder unabhängigen, kontinuierlichen Persönlichkeitsdimensionen zu bestimmen, deren Zusammenhang mit berufsrelevanten Erfolgsfaktoren empirisch nachgewiesen wurde.

Hinsichtlich der Objektivität, genauer gesagt, hinsichtlich der Interpretationsobjektivität unterscheiden sich die am Markt angebotenen Persönlichkeitsfragebogen erheblich. Während die Durchführungs- und Auswertungsobjektivität in der Regel gegeben ist, versagen etliche Verfahren bei der Frage, wie mit den gewonnenen Informationen, z. B. mit dem resultierenden Persönlichkeitsprofil, umzugehen ist. Ein Persönlichkeitsfragebogen, dessen Ergebnis von unterschiedlichen Personen unterschiedlich „gedeutet“ oder „gelesen“ wird, mangelt es an Interpretationsobjektivität. Gerade die bei vielen Anwendern beliebten „freien Interpretationen“ der Ergebnisse von Persönlichkeitsfragebogen sind zu vermeiden. Ein Verfahren mit hoher Interpretationsobjektivität stillt allerdings nicht das Machtbedürfnis mancher Berater. Wenn ein Verfahren objektiv ist, kommen unterschiedliche Berater zum gleichen Ergebnis/zur gleichen Entscheidung. Der Berater ist logischerweise austauschbar. Bei einem Verfahren, dessen Ergebnisse hingegen individuell interpretiert und gedeutet werden, ist der Berater jedoch unentbehrlich, er hat Guru-Status. Der häufige Einsatz von Verfahren mit minderer Qualität (geringe Objektivität und in Folge mangelhafte Zuverlässigkeit und Gültigkeit) dürfte sich wesentlich durch das Machtmotiv einiger Berater erklären. Man kann sich für einzelne Persönlichkeitsfragebogenverfahren als „Master“ oder ähnliches akkreditieren lassen und somit sein Geltungsbedürfnis befriedigen – leider häufig auf Kosten der Qualität der Potentialdiagnose und Personalauswahl.

Ein häufig und heftig diskutiertes Thema im Kontext der Management-/Eignungsdiagnostik mit Persönlichkeitsfragebogen ist deren Verfälschbarkeit. Es ist unumstritten, dass Persönlichkeitsfragebogen verfälscht werden können. Dies gilt ebenso für andere Verfahren der Potentialfeststellung und Personalauswahl wie beispielsweise Interviews oder Assessment Center. Bei Persönlichkeitsfragebogen ist die Verfälschung allerdings besonders leicht möglich. Während die Verfälschbarkeit früher allerdings als ein schwerwiegendes Problem angesehen wurde, welches die Anwendbarkeit von Persönlichkeitsfragebogen zu Auswahlzwecken grundsätzlich in Frage stellte, gibt es inzwischen Hinweise darauf, dass die Verfälschung die Kriteriumsvalidität der Verfahren ggf. nicht grundlegend beeinträchtigt. Überhaupt ist zu fragen, ob die Verfälschbarkeit wirklich ein Problem darstellt oder ob die Fähigkeit zum situationsangemessenen „Faken“ nicht selbst wertvoll für die spätere berufliche Leis-

tung ist (Marcus, 2003). Während Marcus (ebd.) keine negativen Auswirkungen der Verfälschbarkeit auf die Kriteriumsvalidität erkennt, sieht Kersting (2004) u.a. Akzeptanzprobleme für den Einsatz von verfälschbaren Verfahren in der angewandten Managementdiagnostik.

5.5 CUBE Dimension 1–5: Psychometrische Leistungstests (z. B. Tests zur kognitiven Kompetenz)

Bei Persönlichkeitsfragebogen gibt es keine richtigen und falschen Antworten – so wird gerne gesagt. Im Kontext der Personalauswahl stimmt das natürlich nur bedingt: Antworten, die den Vorstellungen der Organisation nicht entsprechen, führen aus der Sicht des Kandidaten zum Bewerbungsmisserfolg und waren so gesehen „falsch“. Dennoch unterscheiden sich die Leistungstests (z.B. Tests zur kognitiven Kompetenz, Wissenstests, Problemlöseszenarien, Konzentrationstests, Sprachtests, Gedächtnistests, case-studies und Postkörbe mit vollstandardisierter Auswertung) im Wesentlichen durch das „scoring“ von Persönlichkeitsfragebogen. Was richtig oder falsch ist, variiert nicht – wie bei der Interpretation eines Persönlichkeitsfragebogens – von Organisation zu Organisation, sondern ergibt sich aus der Aufgabenstellung. Es gibt eine „natürliche“, vom Kandidaten eindeutig erkennbare Bewertung (z.B. bei Tests zur kognitiven Kompetenz und bei Wissenstests nach „richtig“ oder „falsch“ oder bei Tests zur Bearbeitungsgeschwindigkeit nach „schnell“ oder „langsam“). „Vom Kandidaten erkennbar“ meint nicht, dass der Kandidat die Lösung kennt, sondern dass er weiß, dass eine „richtige“ Lösung (eine „richtige“ Bearbeitungsstrategie, z.B. schnell und fehlerfrei) gesucht wird und vorhanden ist – auch wenn er diese nicht findet. Darin liegt eine Stärke der Leistungstests: Die Leistungstests sind hochgradig objektiv, zuverlässig und valide. Die Gültigkeit (Validität) eines Leistungstests zur kognitiven Kompetenz bezüglich der Vorhersage des Berufserfolgs ist beispielsweise mit $r = .66$ (Kramer, 2009) deutlich höher als die Gültigkeit von strukturierten Eignungsinterviews ($r = .44$, McDaniel et al., 1994) und Assessment Centern ($r = .41$, Holzenkamp, Spinath und Höft, 2010). Zugleich liegt in der Objektivität und Transparenz aber auch ein Problem der Leistungstests: Anders als bei einem Interview mit einem Berater, der auch bei leistungsschwachen Kandidaten freundlich und gewinnend agiert, merken die Kandidaten bei einem Leistungstest ggf. selbst, dass Sie den Anforderungen nicht gerecht werden. Zum Selbstschutz werten sie dann unter Umständen das Verfahren ab, der Berater sieht sich der Kritik ausgesetzt, dass das Verfahren unsinnig sei usw. Leistungstests finden bei gestandenen Kandidaten – so die Sorge der Auftraggeber und Berater – keine Akzeptanz. U.a. aus diesem Grunde werden Leistungstests trotz ihrer hervorragenden Aussagekraft außerhalb der Personalauswahl bei Einsteigerpositionen für junge Menschen selten eingesetzt. Dabei ist

die Akzeptanz der Leistungstests nicht so schlecht wie vermutet (siehe Kersting, 2008b) und lässt sich durch verschiedene Maßnahmen (z.B. eine Kontextualisierung der Aufgaben) steigern.

5.6 Besondere Qualitätsanforderungen für psychometrische Fragebogen und Leistungstests (CUBE Dimension 1–4 sowie 1–5)

Im Gegensatz zur Dokumenten- und Faktenanalyse, zur Befragung und zur Verhaltensbeobachtung beruht die Auswertung und Interpretation der meisten psychometrischen Verfahren (CUBE Kategorien 4 und 5) auf einen normativen Ansatz, d.h. man vergleicht zur Interpretation und Bewertung des Verfahrensergebnisses den gemessenen Wert mit den Werten anderer Personen. Daher muss man wissen, welche Vergleichswerte für die Auswertung und Interpretation herangezogen werden können. Gut ist ein psychometrisches Verfahren unter diesem Gesichtspunkt, wenn die Normdaten in jüngerer Zeit an einer sehr großen Gruppe von Personen erhoben wurden, die der Zielgruppe entsprechen. Natürlich müssen den Anwendern umfangreiche Informationen zu den Normgruppen zur Verfügung stehen (z.B. Wann wurden die Normdaten erhoben? Wie setzt sich die Normgruppe zusammen? Haben die Teilnehmer das Verfahren freiwillig und anonym oder im Ernstfall einer Bewerbungssituation bearbeitet?).

Insgesamt sind umfassende Informationen eine notwendige Voraussetzung zur Beurteilung psychometrischer Verfahren (CUBE Kategorien 4 und 5 der ersten Dimension). Nur auf der Basis von Informationen lässt sich angeben, in welchem Umfang die Gütekriterien Objektivität, Zuverlässigkeit und Gültigkeit (sowie weitere Kriterien wie z.B. Ökonomie, Akzeptanz) erfüllt sind. Ein Großteil dieser Informationen besteht notwendigerweise aus Berichten über die Ergebnisse empirischer Studien. Es ist nicht möglich, einen psychometrischen Fragebogen oder Leistungstests allein aufgrund von Aufgaben/Fragen oder anhand von „Muster-Gutachten“ usw. zu beurteilen. Für zahlreiche psychometrischen Fragebogen oder Leistungstests fehlen aber die für eine aussagekräftige Beurteilung notwendigen Informationen. Anstelle der Sachinformationen finden sich blumige Aussagen in Verkaufsprospekten. Welche Informationen konkret zu einem psychometrischen Verfahren vorliegen müssen, hat das Testkuratorium (2010) definiert. Die Kriterien stehen in Form einer Checkliste zur Verfügung, die unter der URL http://kersting-internet.de/DIN-Buch/din-buch_downloads.html kostenlos abgerufen werden kann. Darüber hinaus hat das Testkuratorium ein neues System zur Information über und Beurteilung von psychometrischen Fragebogen oder Leistungstests gestaltet. Alle nach diesem System bislang erstellten Rezensionen können kostenfrei unter der URL <http://www.zpid.de/Testkuratorium> eingesehen werden. Unter dieser URL steht auch der Text des Beurteilungssystems zum

kostenlosen Download bereit. Dieser stellt einen Leitfaden dar, nach dem man die psychometrischen Fragebogen und Leistungstests beurteilen kann, die bislang noch nicht rezensiert wurden.

Psychometrische Fragebogen oder Leistungstests werden nicht mehr nur als „Papier-und-Bleistift“ Verfahren angeboten. Die Bearbeitung kann vielmehr in vielen Fällen entweder offline am stationären PC oder online und somit unabhängig von Ort und Zeit durchgeführt werden. Neben Texten können Bilder, Filme, Audio-Aufnahmen, Simulationen usw. vorgegeben werden, die Darbietungsfolgen können in Abhängigkeit vom Antwortverhalten der getesteten Bewerber variieren usw. (siehe z.B. Bartram & Hambleton, 2006; Kirbach, Montel, Oenning & Wottawa, 2004). Für die Auswertung von Fragebogen und Leistungstests empfiehlt es sich in den meisten Fällen, also auch dann, wenn das Verfahren noch „händisch“ bearbeitet wird, den Computer zu nutzen.

6. Verfahrenspluralismus

Und welches Verfahren ist nun das Beste? In der Managementdiagnostik empfiehlt sich in der Regel eine Kombination unterschiedlicher Verfahren, da mit jedem Verfahren neben den relevanten Informationen auch Eindrücke hervorgerufen werden, die verfahrensspezifisch sind, also mehr von der Eigenart des Verfahrens als von der zu beurteilenden Person abhängen. Die Grundidee des Verfahrenspluralismus (auch Multimethodalität oder Multimodalität) besagt, dass durch die gezielte Kombination unterschiedlicher Verfahren/Datenquellen das interessierende Merkmal, beispielsweise die Leistungsmotivation eines Bewerbers, fokussiert und die nicht intendierten, z.B. verfahrensspezifischen (Varianz-)Anteile unterdrückt werden. Daher sollte in der Praxis eine anforderungsorientierte, theoriegeleitete Kombination unterschiedlicher Verfahren realisiert werden. Mit dem hier vorgestellten CUBE System wird die Verfahrensauswahl erleichtert, indem ein Überblick über alle möglichen Verfahren zur Verfügung gestellt wird und deren Gemeinsamkeiten und Unterschiede transparent werden.

Der Begriff des Verfahrenspluralismus bezieht sich im Rahmen des CUBE Systems nicht nur auf die Nutzung unterschiedlicher Verfahren (Makrostrategie, CUBE Dimension 1), sondern auch auf die vielfältigen Möglichkeiten der Ausgestaltung eines Verfahrens (Mikrostrategie, CUBE Dimensionen 2 und 3). Dies wird deutlich, wenn die Fragen im Interview anhand der Zeitdimension des CUBE in biographische und situative Fragen differenziert werden.

Sofern unterschiedliche Verfahren miteinander kombiniert werden, kommt der Integration der Daten eine entscheidende Bedeutung zu. Zu-

mindest wenn es um Auswahlentscheidungen geht, ist die Bestimmung eines Gesamtwertes notwendig. Alle Fragen der Datenintegration müssen im Vorfeld so eindeutig geregelt werden, dass es unbedeutend ist, wer die Datenintegration vornimmt, da man stets zum gleichen Ergebnis kommt. Anders formuliert: Auch bei einem der letzten Schritte, bei der Datenintegration, kommt es auf Objektivität an.

Literatur

- Arbeitskreis Assessment Center (AK-AC) (2004). *Standards der Assessment Center Technik*. Hamburg: Autor.
- Bartram, D. & Hambleton, R. (Eds.). (2006). *Computer-based testing and the internet: Issues and advances*. New York: John Wiley.
- Brannick, M. T., Levine, E. L. (2002). *Job analysis: Methods, research, and applications for human resource management in the new millenium*. Thousand Oaks: Sage Publications.
- DIN. (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.
- Holzenkamp, M., Spinath, F. M. & Höft, S. (2010). Wie valide sind Assessment Center im deutschsprachigen Raum? Eine Überblicksstudie mit Empfehlungen für die AC-Praxis. *Wirtschaftspsychologie*, 12, 17–25.
- Hossiep, R. & Mühlhaus, O. (2005). *Personalauswahl und -entwicklung mit Persönlichkeitstests*. Göttingen: Hogrefe.
- Hossiep, R. & Paschen, M. (2003). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP) (2. Aufl.)*. Göttingen: Hogrefe.
- Jäger, R. S. (2004). Test im Test. Insights MDI – Wissenschaftlich betrachtet. *PersonalMagazin*, 1, 22.
- Kersting, M. (2004). Zur Bedeutung der Validität und der sozialen Akzeptanz in der Berufseignungsdiagnostik. *Zeitschrift für Personalpsychologie*, 3, 83–86.
- Kersting, M. (2006). Stand, Herausforderungen und Perspektiven der Managementdiagnostik. *Personalführung*, 10, 16–27.
- Kersting, M. (2008a). *Qualität in der Diagnostik und Personalauswahl: Der DIN Ansatz*. Göttingen: Hogrefe.
- Kersting, M. (2008b). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420–433.
- Kersting, M. (2008c). Zur Zukunft von Assessment-Center-Verfahren. Möglichkeiten der Qualitätssicherung und -optimierung. *Personalführung*, 12/2008, 72–80.
- Kersting, M. (2010). Personalauswahl: Methoden für erfolgreiche Unternehmen. In: Berufsverband Deutscher Psychologinnen und Psychologen (Hrsg.), *Psychologische Expertise für erfolgreiches Unternehmertum in Deutschland*, S. 57–65. Berlin: Deutscher Psychologen Verlag.

- Kirbach, C., Montel, C., Oenning, S. & Wottawa, H. (2004). *Recruiting und Assessment im Internet. Werkzeuge für eine optimierte Personalauswahl und Potenzialerkennung*. Göttingen: Vandenhoeck & Ruprecht.
- Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: Vertiefende und weiterführende Metaanalysen. *Psychologische Rundschau*, 60, 82–98.
- Marcus, B. (2003). Persönlichkeitstests in der Personalauswahl: Sind „sozial erwünschte“ Antworten wirklich nicht wünschenswert? *Zeitschrift für Psychologie*, 211, 138–148.
- 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.
- Sarges, W. (2006). Management-Diagnostik. In: F. Petermann & M. Eid (Hrsg.). *Psychologische Diagnostik*, S. 739–746. Göttingen: Hogrefe.
- Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
- Schuler, H. (2006). Arbeits- und Anforderungsanalyse. In: H. Schuler (Hrsg.) *Lehrbuch der Personalpsychologie* (2. Aufl., 45–68). Göttingen: Hogrefe.
- Schuler, H. & Prochaska, M. (2000). Entwicklung und Konstruktvalidierung eines berufsbezogenen Leistungsmotivationsstests. *Diagnostica*, 46, 61–72.
- SHL (2000). *OPQ32. Manual and User's Guide*. Thames Ditton: Saville & Holdsworth.
- Spinath, F. M. & Becker, N. (2011). Verhaltensbeobachtung. In: L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich B, Methodologie und Methoden. Serie II, Psychologische Diagnostik. Band 2, Methoden der psychologischen Diagnostik*, 325–369. Göttingen: Hogrefe.
- Westhoff, K. & Strobel, A. (2011). Interview. In: L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich B, Methodologie und Methoden. Serie II, Psychologische Diagnostik. Band 2, Methoden der psychologischen Diagnostik*, 371–413. Göttingen: Hogrefe.
- Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen (2010). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 09. September 2009. *Psychologische Rundschau*, 61, 52–56.
- Wernimont, P. F. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372–376.